CHANGES IN SOCIAL NETWORK STRUCTURE IN RESPONSE TO EXPOSURE TO FORMAL CREDIT MARKETS

ABHIJIT BANERJEE[†], EMILY BREZA[#], ARUN G. CHANDRASEKHAR[‡], ESTHER DUFLO[§], MATTHEW O. JACKSON[⋆], AND CYNTHIA KINNAN[¶]

Abstract. We show that the entry of formal financial institutions can have far-reaching and long-lasting impacts on informal lending and social networks more generally. We first study the introduction of microfinance in 75 villages in Karnataka, India, 43 of which were exposed to microfinance. Using difference-in-differences, we show that networks shrank more in exposed villages. Moreover, links between households that were both unlikely to borrow from microfinance were at least as likely to disappear as links involving likely borrowers. We replicate these surprising findings in the context of a RCT in Hyderabad, where a microfinance institution randomly selected 52 of 104 neighborhoods to enter first. Four years after all neighborhoods were treated, households in early-entry neighborhoods had had credit access longer and had larger loans. We again find fewer social relationships between households in these neighborhoods, even among those ex-ante unlikely to borrow. Because the results suggest global spillovers, atypical in usual models of network formation, we develop a new dynamic model of network formation that emphasizes chance meetings, where efforts to socialize generate a global network-level externality. Finally, we analyze informal borrowing and the sensitivity of consumption to income fluctuations. Households unlikely to take up microcredit suffer the greatest loss of informal borrowing and risk sharing, underscoring the global nature of the externality.

JEL Classification Codes: D85, D13, L14, O12, Z13

Keywords: Social Networks, Network Change, Network Formation, Network Evolution, Microfinance, Market Exposure, Favor Exchange, Social Capital

Date: February 28, 2022.

We thank Siwan Anderson, Patrick Francois, Rachel Kranton, and participants at various seminars for suggestions. Financial support from the NSF under grants SES-1156182, SES-1155302, SES-1629446, and SES-2018554; and from the AFOSR and DARPA under grant FA9550-12-1-0411; and from ARO MURI under award No. W911NF-12-1-0509 is gratefully acknowledged. We thank Devika Lakhote, Anirudh Sankar, Varun Kapoor, Stephen Nei, Mounu Prem, Tristan Loisel and Sumhith Aradhyula as well as the CMF at IFMR, Tanay Balantrapu, Gowri Nagraj, and Manaswini Rao for excellent research assistance and helpful discussions.

[†]Department of Economics, MIT; NBER; JPAL.

[#]Department of Economics, Harvard University; NBER; JPAL.

[‡]Department of Economics, Stanford University; NBER; JPAL.

[§]Department of Economics, MIT; NBER; JPAL.

[⋆]Department of Economics, Stanford University; Santa Fe Institute.

[¶]Department of Economics, Tufts University; NBER; JPAL.

1. Introduction

Social networks are an important source of credit, insurance, information, advice, and other economic and non-economic benefits substituting for absent or poorly-performing formal institutions.[1] But social networks are not designed: they emerge as the product of many decentralized decisions. In particular, as formal markets expand, the incentives to maintain or develop new relationships change. This could affect networks in unanticipated ways, potentially affecting even those who do not directly benefit from this expansion (Arrow, 2000; Putnam, 2000).

In this paper, we study how the introduction of formal lending institutions changes social networks, both empirically and theoretically. In our first empirical setting, we analyze how the introduction of microfinance (MF) affects network relationships in rural communities. We show that MF entry leads to a general reduction in network links, including among those whose characteristics make them very unlikely to be borrowing from the microfinance institution (MFI). In fact, despite being *prima facie* unlikely to be involved with microcredit, they are at least as affected, and sometimes more affected than those who join microcredit. In particular, their relationships with others who, like them, are unlikely to join microcredit shrink considerably. Because existing models of network formation struggle to rationalize these patterns, we develop a new model that can explain these findings. Our model highlights spillovers stemming from the decision to socialize or not. We subsequently replicate these surprising findings in a second, independent empirical setting where randomly chosen urban communities get access to microcredit, demonstrating the robustness of these findings. Moreover, in this case we are able to show that the loss in links persists even after microfinance is no longer available to these communities.

The challenge in ascertaining whether formal institutions change informal social structures is that it requires detailed data on networks of informal relationships and exogenous variation in access to formal institutions. Our two empirical contexts satisfy both requirements. First, we analyze the introduction of MF in rural Karnataka, India using two waves of detailed network panel data that we collected (Banerjee, Chandrasekhar, Duflo, and Jackson, 2013, 2019b) over six years in 75 villages. These villages were selected in 2006, prior to the first survey wave, when none of them had access to microfinance, but a microfinance institution, Bharatha Swamukti Samsthe (BSS), was planning to start operating in all of them. Between 2007 and 2010, BSS entered 43 of these 75 villages, which we call MF villages. However, a series of external crises halted BSS's expansion and the remaining 32 villages were not exposed to BSS prior to our Wave 2 survey, collected in 2012. We call these non-MF villages. We take advantage of this variation, along with our extremely detailed network data from the two waves (covering 16,476

[1]See, e.g., Udry (1994); Fafchamps and Lund (2003); Karlan, Mobius, Rosenblat, and Szeidl (2009); Beaman and Magruder (2012); Ambrus, Mobius, and Szeidl (2014); Blumenstock, Eagle, and Fafchamps (2016); Munshi and Rosenzweig (2016); Blumenstock and Tan (2016); Breza (2016).

households) to estimate the impact of MF on village network structure using a difference-in-difference strategy.

Second, we replicate and extend the Karnataka findings, leveraging an RCT conducted in 104 neighborhoods in Hyderabad, India (Banerjee, Duflo, Glennerster, and Kinnan, 2015a; Banerjee, Breza, Duflo, and Kinnan, 2019a). In the RCT, entry by an MFI (Spandana) was randomized to half of the study neighborhoods. Control areas began receiving access to Spandana two years later. But in 2010 Spandana suddenly ceased all operations, due to the same set of crises that halted BSS's expansion. We surveyed all households six years after initial entry of microcredit. At this point they had little or no access to microcredit, though households in the early entry neighborhoods had been exposed for twice as long before microcredit was shut down and had received larger loans. We estimate the impact of this differential access to microcredit.

The advantage of the Karnataka setting is that we have very high-quality network data. We know details of link patterns between households as well as the nature of the link (e.g., financial, informational, social). Furthermore, since we have panel network data, we can condition on pre-period network structure. However, the setting does not involve an RCT and therefore our identification relies on the difference-in-difference estimator being valid.

The Hyderabad dataset avoids this issue, since initial entry was randomized and, as a result, treatment neighborhoods had exogenously more cumulative access to microfinance than control neighborhoods. Also, because the survey was fielded 6 years after initial entry and 4 years after the late-entry group received access to MF, the results indicate that these kinds of effects can be durable. Finally, the hypotheses we test in this data come from the results of the Karnataka analysis, which were generated before we looked at the network data in Hyderabad. In this sense these results have the potential to validate the takeaways from Karnataka. However the Hyderabad network data is more limited than the Karnataka data–we only have one cross-section of network information and only partial network data. To supplement it, we collected "aggregated relational data" (ARD) and use the new methodology from Breza et al. (2020b, 2019) to estimate features of the network. Our ARD survey asks each respondent to list their network relationships and to indicate how many of those individuals have a series of traits (e.g., a household member who migrated abroad, a government job). Breza, Chandrasekhar, McCormick, and Pan (2020b) and Breza, Chandrasekhar, McCormick, and Pan (2019) have shown that these responses contain sufficient information to identify the parameters of a network formation model which can then be used to estimate the key characteristics of the neighborhood network that we need for our analysis. Breza et al. (2020b) and Breza et al. (2019) show that this method is an effective way of identifying effects on networks, with very little loss compared to the case where the researcher has full network data.

The impact of microfinance on network connections for those involved can potentially go in either direction. As a source of formal credit to poor, underbanked households, microfinance may reduce dependence on social networks for informal credit and insurance. Moreover, the required weekly repayment structure of microloans may reduce borrowers' liquidity and limit their capacity to lend small sums to their friends (Field, Pande, Papp, and Park, 2012). On the other hand, if households re-lend a part of their microfinance loans, microfinance could crowd in informal financial relationships.[2]

In both of our datasets we find that the introduction of microfinance crowds out social network relationships. The probability of a link between any two households declines by 11% ($p = 0.077$) in a MF village compared to a non-MF village in the Karnataka sample. This is robust to controlling for a rich array of baseline variables. We estimate an even larger effect in the Hyderabad RCT – a 22% decline ($p = 0.062$).

We then investigate how the changes in networks are distributed across two types of households: those who are likely to take up microfinance loans and those who are unlikely to do so. All of the channels described above suggest that microfinance might affect borrowers' willingness to maintain friendships, including with those who do not take up microfinance. However, *prima facie* (without any sort of externality or spillover), one would not expect effects on pairs (or groups) of households that are *both* unlikely to take up microfinance. If anything, one would have expected links between these households to be *strengthened* in microfinance villages, since they might be losing access to the households that get microfinance but still have needs to borrow and lend.

To look at this question empirically, we need to be able to compare those who are more vs. less likely to take up microfinance in MF villages/neighborhoods to those in a non-MF village who would have been comparably likely to take up microfinance had it been available in their village/neighborhood. To this end, we use a random forest model to classify households in all villages into two groups based on whether they would have a high (H) or low (L) likelihood of joining microfinance if it were offered in their village.[3]

Our empirical analysis focuses on both links and triangles (three nodes mutually linked), the latter capturing impacts on local group structures which are known to play a major role in sustaining informal insurance (Bloch et al., 2008; Jackson et al., 2012). There are three link types—LL, LH, and HH—and four triangle types—LLL, LLH, LHH, HHH.

Our main contribution is to show that, across the board in both datasets, links and triangles among those uninvolved with microcredit (LL, LLL) decline as much as links and triangles involving H-types. This is inconsistent with the intuition, as suggested above, that those not exposed at all should respond less than those partially or directly exposed. It is also inconsistent with accompanying models we describe in Appendix D, which largely predict that we should see exposure effects that parallel the direct impact of treatment.

[2]Kinnan and Townsend (2012); Field, Pande, Papp, and Park (2012); Feigenberg, Field, and Pande (2013) and Vera-Cossio (2019) find evidence consistent with the households re-lending bank and credit cooperative loans.

[3]The random forest classifier performs at least as well as a conventional logistic regression-based classifier, and strictly better in one dataset, suggesting there may be a high degree of non-linearity in the prediction function.

Specifically, in the Karnataka panel, when we examine the probability that two Ls who were linked in Wave 1 (LL links) continue to be linked in Wave 2, we obtain the surprising result that LL links decline as much as LH links and more than HH links in MF relative to non-MF villages. An LL link that exists in Wave 1 in a MF village is 5.8pp ($p = 0.002$) less likely to exist in Wave 2 compared to a similar link in a non-MF village; the p value for the difference in coefficients between LL and HH links is 0.086 without controls and 0.292 with controls. Similarly, new LL links are less likely to form in MF villages compared to new HH links (the p-value is 0.206 without controls and 0.059 with controls).

The cross-sectional network data from the Hyderabad RCT delivers consistent results. Treated MF neighborhoods have 0.7pp (22%) fewer LL links than control neighborhoods ($p = 0.004$), and there is no evidence of a greater treatment effect on LH or HH links.

We then examine the evolution of links that form triangles. In the Karnataka sample, we find that it is the LLL triangles that are most likely to disappear in MF villages compared to non-MF villages. In MF villages, LLL triangles are 7.8pp ($p = 0.008$) more likely to have at least one link broken than in non-MF villages, more than any other type. The difference is greatest and most significant between LLL and HHH, but even LHH are less likely to break than LLL (by 5.4pp, $p = 0.072$). LLL triangles are also more likely to entirely disappear in MF villages, and the difference from all of the other types of triangles is significant. In the Hyderabad data, we also find that we are significantly less likely to observe a LLL triangle in treatment than control villages.

It is instructive that in one of the two contexts, LLs and LLLs decline more than their counterparts. As a back-of-the-envelope calculation, if we look at all pairwise comparisons of LL versus its counterparts (HL, HH) and also LLL versus its counterparts (HHL, HLL, HHH) and focus on the most stringent regressions with myriad controls in the main body of the paper, 35.7% of the parameters are significant at the 10% level. This is a crude but useful way to guard against false discovery, since that number should be 10% under the null of no effect. If we turn to our robustness exercises for classification in Appendix F, that number climbs to 50%. The evidence suggests that in Karnataka, LLs and LLLs are dropping at faster rates than their counterparts.

In Hyderabad, we find instead that LL links and LLL triangles drop *as much* as groupings involving Hs. However, the fact that we see *greater* drops in Karnataka means that whatever dynamics underpin the behavioral response by individuals must allow for these "non-monotone" effects.

Moreover, the fact that LLs and LLLs drop comparably to their counterparts suggests the presence of a large externality mediated by something other than mutual consent. Further, there is yet another externality that we identify: one that operates across multiplexed layers of relationships. In Karnataka, even though the direct impact of microfinance is likely to be on financial links, the same patterns also emerge when we analyze information (i.e., advice-giving and -receiving) links. In Hyderabad, the evidence for this phenomenon is more suggestive among Ls, but a similar multiplexing spillover is seen for the Hs. Overall, this suggests that there is contagion from one type of relationship to others.

These types of spillovers, both across types of links and across types of households, are *prima facie* inconsistent with models of network formation where the decision to form a link only depends on the payoff to the two parties forming the link and where these payoffs only depend on the characteristics of the two parties involved in the link and no one else. We briefly sketch a set of these models that are standard in the literature in Online Appendix D: these include models of directed search with mutual consent needed to form links, stickiness in dropping or forming links, and local payoff externalities.[4]

Based on the empirical results from the Karnataka dataset, and prior to analyzing the Hyderabad data, we developed a new model of network formation that can explain why links between the Ls might break as much as other links, and possibly even more. The model has two types of externalities: (1) individuals must mutually consent to linking with other candidates they meet; (2) individuals must engage in the socializing process in the first place. The first externality alone is inconsistent with the patterns of the data, but the second mechanism delivers the results that Ls can lose at least as much as Hs. In the model, old relationships are maintained and new ones are formed when people socialize in an "undirected" way. A stylized interpretation is that people show up at the town square, or a local tea shop, to "hang out" and socialize. Seeing their current friends keeps those relationships intact, and meeting new people sometimes results in new relationships. People who do not show up at the town square lose old relationships and form fewer new ones. We describe this as a model of undirected search.

This gives rise to a distinct *network-level* externality, because the returns to socializing depend on who else is socializing. Holding fixed the valuation of a certain link or groups of links, the fact that, in equilibrium, others are not searching can have global effects on network density and topological structure. For example, L types value HL links, and thus care about how H types socialize. Therefore, if microfinance changes the socialization of H types, the incentives for L types to socialize also change, which in turn affects the incidence of LL links. Specifically, access to microcredit might reduce both the demand and supply of informal loans by H types, but the Hs becoming less willing to lend can have a larger negative impact on Ls than on Hs, which leads to less socializing by Ls. As Ls socialize less there is a larger relative drop in LL links. A simple extension of the model to account for the formation of triads (triangles) generates similar results for LLL relationships. This model matches the patterns we observe in the data, in particular the spillovers onto the relationships between L types. It also predicts

[4]See, e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami (1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den Nouweland (2005); Herings et al. (2009); Boucher (2015); Watts (2001); Jackson and Watts (2002); Christakis et al. (2010a); König et al. (2014); Currarini et al. (2009, 2010); Cabrales et al. (2011); Canen et al. (2017).

that there should be spillovers across different types of relationships, since it is the same town square where people also form other kinds of relationships.

Given the loss of links by Ls, a natural question is whether we see changes in downstream outcomes, such as borrowing or the volatility of consumption. Consistent with the disappearance of the LL links, we find in both settings that the L households, after the introduction of microfinance, borrow relatively less from informal sources in MF compared to non-MF villages.

Finally, in the Hyderabad sample we can directly measure the impact of increased microfinance exposure on consumption smoothing for H vs. L households. This is possible because we have detailed household-level panel information on both income and consumption. In areas exposed to microfinance, households with high propensity to use microfinance (Hs) see little change in their consumption smoothing compared to those in areas not exposed to microfinance. However, households with low propensity to use microfinance (Ls) see a large and significant worsening of their consumption smoothing compared to those in areas not exposed to microfinance, which is consistent with the network and informal borrowing impacts.

Our research on how exposure to formal financial institutions affects social and economic networks is related to several bodies of work. Several strands of literature have explored the interplay between formal and informal institutions in the context of economic development. Coate and Ravallion (1993) note that informal insurance constrained by limited commitment can arise as a substitute for formal insurance. Kranton (1996) highlights the potential for market-based exchange to crowd out reciprocal exchange. Ligon et al. (2000) highlight the potential for access to savings to crowd out interpersonal insurance. These papers are theoretical in nature; in addition, a number of papers have studied this interplay empirically. For instance, Albarran and Attanasio (2003) show that access to a cash transfer program crowds out private transfers in Mexico, while McMillan and Woodruff (1999) show that the absence of legal enforcement of contracts sustains informal firm to firm lending in Vietnam, and Macchiavello and Morjaria (2021) show that competition crowds out relational contracts. Our paper builds on these literatures by showing what is, to our knowledge, the first empirical evidence showing that changes in access to a formal institution (microfinance) can cause the deterioration of informal social networks, not only for those adopting microfinance, but also among those unlikely to adopt microfinance and also affect network relationships that do not involve any material exchanges.

Our findings are also directly related to recent literature exploring the impacts of access to formal financial services on social networks. Feigenberg et al. (2013) find that participation in microcredit creates tighter social relationships among participating group members. Binzel, Field, and Pande (2013) and Comola and Prina (2017) explore whether and in what ways financial interventions affect participating households' networks.[5]

[5]Specifically, Binzel et al. (2013) look at network effects in a randomized roll-out of branches of a new financial intermediary in India. Their focus is on whether individuals are less likely to make transfers to their friends in a non-anonymous dictator game after being exposed to the financial institution. Comola and Prina (2017) study

In recent work, Heß, Jaimovich, and Schündeln (2020) also examine how policy interventions affect network structure, but in the context of a community-driven development initiative (CDD). The initiative provided a very large disbursement—one half of annual per capita income *per household* in each treatment village — and villagers had to collectively decide which projects to execute. Heß et al. (2020), like us, document declines in network density and closure, which in their case are generated by political maneuvering and elite capture. Two key differences between CDD and microcredit are that the former involves much larger sums of money and much more coordination at the community level, both of which were probably a source of conflict in Heß et al. (2020) (absent in our setting) and might have caused the changes in the networks. In this sense what we document is a quite different kind of fragility in networks.

The main lesson from our paper is that significant and widespread spillovers in network formation are present across types of people and types of relationships, which is indicative of a global network externality. We also use this evidence to build and argue for a new model of network formation that highlights the fact that social networks are not "designed" but result from the decentralized decisions of individuals. As our empirical results highlight, in such an environment, a shift in the incentives of one group of people to form links can have substantial negative effects on other groups in the network that the first group ignores when choosing their own behavior.[6] Of course, this does not directly imply that microcredit should be discouraged, but rather that any welfare analysis needs to take into account the potential for spillovers on non-adopters.

The remainder of the paper is organized as follows. In Section 2, we describe the setting, network data collection, the classification of households into H and L types using a random forest algorithm, and sample statistics. In Section 3, we present our empirical results. Motivated by the Karnataka data, in Section 4 we develop a new dynamic model of network formation that is consistent with our findings and discuss why four standard models from the literature are inconsistent with the data. In Section 5, we present impacts on informal borrowing and the capacity for households to smooth consumption. In Section 6, we conclude.

2. Setting, Data and Sample Statistics

2.1. **Setting.**

2.1.1. *Karnataka (India).* In 2006, the microfinance organization, BSS, provided us with a list of 75 villages in Karnataka in which they were planning to start lending operations. The villages were spread across 5 districts of the state of Karnataka in India. Prior to BSS's entry, these villages had minimal exposure to microfinance.

---

spillovers due to the randomized introduction of savings accounts in Nepal. They find that those randomly given accounts were less likely to lose links present at baseline and more likely to add links to untreated households. Relatedly, Dupas et al. (2019) show that access to savings accounts crowded in interpersonal transfers in Kenya.

[6]See Jackson (2003) for background on inefficiencies in network formation.

Six months prior to BSS's entry into any village, in 2006, we conducted a baseline survey in all 75 villages. This survey consisted of a village questionnaire, a full census that collected data on all households in the villages, and a detailed follow-up survey fielded to a subsample of adults.

By the end of 2010, BSS had entered 43 villages that were not randomly assigned by us, but rather selected by the bank. We have anecdotal reasons to believe that the choice was not systematic: BSS planned to enter all of the villages but slowed down and ultimately stopped expanding during the Andhra Pradesh (AP) microcredit crisis (see Breza and Kinnan (2021) for background on that crisis).

2.1.2. *Hyderabad (India).* In 2006 Spandana–a large microfinance institution–randomly chose 52 of 104 neighborhoods in Hyderabad (at the time the capital of Andhra Pradesh, a State neighborhing Karnataka, in South India) to enter. After two years, the remaining 52 neighborhoods received access in mid-2008. The short- and medium-run impacts of randomized access to microfinance in this setting are studied in Banerjee et al. (2015a). The AP microcredit crisis also impacted Spandana and its lending activities in Hyderabad. In 2010, all of the households in the Hyderabad sample faced simultaneous withdrawal of microcredit in response to an ordinance halting microcredit loans (this also means they did not need to repay existing debt).[7] A third round of data collection was done in 2012, with a sample of 5744 households. At the time of the 2012 data collection, the treatment neighborhoods had been exposed to microcredit for 6 years (4 years of active lending) and the control neighborhoods had been exposed for 3.5 years (1.5 years of active lending). Network data was collected during this third round.

The early treatment neighborhoods had greater microfinance access overall. Because microfinance borrowers typically receive larger loans each time they borrow, microcredit supply is increasing in the length of exposure. Banerjee et al. (2019a) show that two years after the control group received access, households in treated neighborhoods still had 14% more contemporaneous microfinance borrowing and 43% more cumulative microfinance borrowing over the preceding three years (Banerjee et al., 2019a). However, since nobody had access to microfinance at the time of our network survey, any changes to network structure that we pick up must be the result of the extra exposure to microcredit before it was shut down some two years before our survey. In other words, the effect persists despite there being no differences in contemporaneous participation in microcredit.

2.2. **Data.**

[7]The treatment areas had slightly higher loan balances in 2010 and therefore received a marginally larger windfall associated with the default of existing loans. See Section 5 of Banerjee et al. (2019a) for evidence documenting that this does not explain the differences seen in 2012.

2.2.1. *Karnataka.* To collect the network data,[8] we asked adults to name those with whom they interact in the course of daily activities. In Wave 1, collected in 2006, we have the full village census (enumerating every individual in every household in every village and some basic household characteristics) and network data from 46% of households per village. In Wave 2, collected in 2012, in addition to taking the full village census again, we have network data from 89.14% of the 16,476 households. This means that we have network data in Wave 1 on 70.8% of the links and in Wave 2 on 98.8% of the links when we build the undirected, unweighted graph that we study.[9] For the network analysis, we concentrate on households that are present in both waves and only look at objects (e.g., potential links or potential triads) where we are able to discern in both waves whether the structure exists or does not exist.

We have data about 12 different types of interactions for a given survey respondent: (1) whose houses he or she visits, (2) who visits his or her house, (3) relatives they socialize with, (4) non-relatives they socialize with, (5) who gives him or her medical help, (6) from whom he or she borrows money, (7) to whom he or she lends money, (8) from whom he or she borrows material goods (e.g., kerosene, rice), (9) to whom he or she lends material goods, (10) from whom he or she gets important advice, (11) to whom he or she gives advice, (12) with whom he or she goes to pray (e.g., at a temple, church or mosque).

Using these data, we first look at the financial network (a union of (6-9) above) as well as the informational network ((10-11) from above). After demonstrating that links across both categories change in similar ways, we aggregate the network data as follows. We construct one network for each village, at the household level, where a link exists between households if any member of either household is linked to any other member of the other household in at least one of the 12 ways. We assume that individuals can communicate if they interact in any of the 12 ways, so this is the network of potential communications. The resulting objects are undirected, unweighted networks at the household level.

We also asked, in both Wave 1 and Wave 2, for households to give us a list of all outstanding loans that they have taken, the sources of these loans (e.g., family member, friend, microfinance institution, self-help group, money lender) and their terms. We use this to create a panel to study changes in borrowing patterns.

In our analysis we look at all households who existed in Wave 1 (and in Wave 2 as well). This involves those who remained and those who split. We match households who split in Wave 2 to their Wave 1 counterpart. 11% migrated out, though this is not differential by microfinance exposure, and 4.8% Wave 2 households in-migrated (which we cannot use in the

[8]The Wave 1 data are described in detail in Banerjee, Chandrasekhar, Duflo, and Jackson (2013) and publicly available at `http://economics.mit.edu/faculty/eduflo/social`. The Wave 2 data will be available upon publication.

[9]The 70.8% figure is calculated as follows. Because we consider a non-directed graph, we learn about the existence of a link when either participating node is sampled. Therefore for arbitrary nodes A and B, Pr(sample either A or B) $= 1 - (1 - 0.46)^2 = 0.708$.

panel) or split off from existing households (as children reach adulthood), again not differential by microfinance exposure.[10]

2.2.2. *Hyderabad.* The Hyderabad analysis draws on three waves of data. These data are also utilized in Banerjee et al. (2015a) and Banerjee et al. (2019a). The first round of data collection was conducted in late 2007 - early 2008, 15-18 months after microfinance was made available in the treatment group. Following this first wave, the control group also received access to microfinance in May 2008. A second round of data collection was conducted in mid-2010 to examine longer-term impacts of access to microfinance; coincidentally, this wave took place just before the AP Crisis, mentioned above. Finally, in 2012, approximately two years after the AP Crisis, a third wave of data collection took place. All three waves collected information about household composition, income, consumption/expenditure, borrowing (from microfinance and from other sources) and entrepreneurship.

For the third wave only, we also measured aspects of households' social networks. However, as we could only collect partial network data across the 104 neighborhoods in Hyderabad, we chose to collect Aggregated Relational Data (ARD), which is used as described below. Because we collected this information only in the 2012 Wave 3 data, the majority of our analysis uses Wave 3 only; an exception is the analysis of consumption smoothing which leverages the panel nature of the consumption data.

Specifically, an average of 55 households in every neighborhood in the Hyderabad sample were surveyed and asked a set of network questions. First, respondents were asked how many links they had within the neighborhood (eliciting their degree) along three dimensions: financial, social and informational.[11] This is the directly solicited part of the network information. Second, respondents were asked 9 ARD questions of the form "How many individuals from your neighborhood do you know who have trait X?" For instance, traits include "How many other households do you know where there are 5 or more children?" and "How many other households do you know where any member is a permanent government employee?" Supplemental Appendix E.1 details both types of survey questions. Third, we asked each sampled household whether they themselves possessed each of the ARD traits.

We follow Breza et al. (2020b) in leveraging our ARD survey to estimate key network characteristics. Conceptually, the approach is to observe that if an individual i knows many individuals with trait A but few with trait B, i must be more likely to be linked to As than Bs. Further, if individual i is of type C, then this means that Cs are more likely to link to As than to Bs. In that sense, ARD allows the researcher, under appropriate parametric assumptions, to model

[10]Note that when we construct the panel, our sample of potential links ij conditions on the event that either i or j was surveyed in period 1 in the case of links (and the analogous construction for triangles). Thus, we can be sure that we are studying the evolution of links or triangles in a way that is not plagued by sampling issues (Chandrasekhar and Lewis, 2014).

[11]Specifically, we asked who individuals would go to and who would come to them for borrowing basic goods (cooking gas, a small amount of cash, etc.), advice (e.g., on health or education), and socializing (watching TV).

the underlying network structure despite not having collected link-level data, which can then be used in econometric estimation. Breza et al. (2020b) develop this method, and we also provide further details in Supplemental Appendix E.2.

A rough sketch is as follows. We assume a standard parametric model of network formation (Hoff, Raftery, and Handcock, 2002). There is some (unobserved) latent space where households reside and every household has some location z_i as well as a fixed effect ν_i, which captures an unconditional shifter to their rate of linking. We can think of the probability of nodes linking as proportional to $\exp(\nu_i + \nu_j - \text{dist}(z_i, z_j))$, where $\text{dist}(z_i, z_j)$ is the distance between the two households i and j in this unobserved space.

The location of households in this latent space is modeled as follows. For simplicity of explanation, assume there are K types of households. Every type k has a type center (ζ_k) in the space with some associated variance so that a household i of type k has a location distributed $z_i \sim F(z; \zeta_k, \sigma_k^2)$. In other words, households are more likely to reside close to their type centers but are randomly distributed around this center.

If we have ARD, i.e., information about the number of links i has to members of type k, then Breza et al. (2020b, 2019) show that all type centers and variances, $\zeta_{1:K}$, $\sigma^2_{1:K}$, and all household fixed effects and locations, $\nu_{1:n}$, $z_{1:n}$, can be consistently estimated. That is, all the parameters of the network formation model can be recovered.

The intuition is as follows. For example, if individuals who know families with international migrants also tend to know families with government employees, then international migrants and government employees are likely to be located near one another on the latent space. Moreover, the heterogeneity in friendship patterns across trait groups pins down whether a trait group is tightly concentrated or not. If all individuals with international migrants tend to have similar linking rates to members of their own community, then there is not going to be much dispersion of member locations in the latent space. In this way, the ARD approach allows us to identify trait group locations as well as locations of individual households and their fixed effects.

This is useful because once the Bayesian estimation procedure of Breza et al. (2020b) produces a posterior distribution of the model parameters, we can generate a distribution over the unobserved graph. For each graph realization consistent with that distribution, we can compute the network statistics of interest, such as the linking probabilities for all potential pairs and triples. Finally, we take the expectation of the graph statistic of interest across a large number of realizations. For many applications, this type of network information is enough to draw relevant conclusions.[12] Note that the way we elicited the ARD means that we only

[12]For example, Breza et al. (2020b) show that using ARD, it is possible to replicate the results in Breza and Chandrasekhar (2019) comparably well as when the entire network is observed. We also validate ARD in the Hyderabad dataset. Specifically, in the surveys, we directly measured support – the likelihood that for any link, there exists a third person who has a relationship with both nodes. We validate ARD by comparing the estimated measure of support using the ARD algorithm with the directly elicited survey measure and show that the ARD estimate leads to very similar conclusions.

have information about one single type of link encompassing all dimensions of interaction, both financial and non-financial.

2.3. **Sample Statistics and Covariate Balance.** Starting with Karnataka, Table 1, Panel A shows Wave 1 household demographics by treatment status. Only one covariate out of 20, household size, is statistically different across treatment arms at the 5% level. Appendix Table C.1 includes Wave 1 network characteristics. We find that MF villages are larger, on average, than non-MF villages, likely because the MFI expanded outwards from Bangalore. This in turn results in other differences. However, we show that, conditional on village size, all of the key baseline network characteristics are balanced.[13]

In Appendix Table C.1, we observe that the Wave 1 networks are sparse: the average density is 11.9%. The average clustering coefficient (the percent of cases where two of a household's friends are themselves friends) is 0.33. Finally, these networks have short distances: the average closeness (the mean of the inverse of path lengths, with nodes in different components assigned 0) is 0.379.[14] We present summary statistics for the non-microfinance villages in Wave 2 in Online Appendix Table C.2, Panel A.

We next turn to Hyderabad. Table 1, Panel B shows baseline neighborhood characteristics and pre-determined household demographics by treatment status. As expected, given that the introduction of microfinance was randomized, the covariates are balanced in treatment and control.

Recall that in Hyderabad we have only endline cross-sectional network data, so we only measure the network characteristics after the intervention, and therefore cannot test for baseline balance. In Online Appendix Table C.2, Panel B, we show the means of network characteristics in control neighborhoods. In this urban sample, the networks are even more sparse than in Karnataka; the average degree is 6.0, for an average neighborhood size of approximately 200 households. Average clustering and closeness are also smaller than in Karnataka.

2.4. **Classifying Nodes as H and L.** In order to study heterogeneity in effects by propensity to participate in microfinance, we need to identify which households *would have* taken out microfinance loans in the non-microfinance villages or neighborhoods, had BSS or Spandana entered those locations. To do this, we use a random forest model to classify an individual's propensity to take up microfinance as a function of baseline characteristics, in the presence of

[13]In Online Appendix Tables **??** and **??**, we show that our main results are robust to allowing for differential trends by functions of village size interacted with link or pair type.

[14]In order to deal with the fact that we sampled data in Wave 1, we compute average density among the sampled households in Wave 1, comparing the share of realized links relative to potential links when we fully observe the potential link. We compute the clustering coefficient among the subgraph induced by restricting to sampled households in Wave 1, since that is centered around the true parameter. It is also worth noting that the correlation among the different link types (specifically multiplexing of information and financial links) is 0.638.

microfinance. We can then use this classification exercise to predict which individuals in the entire sample (treatment and control) have a high propensity to borrow from MFIs.

We begin with the Karnataka setting. One obvious determinant of microfinance take-up is based on BSS's rules: only households with a female in the age range 18-57 were eligible for microfinance. Also, certain households were identified by BSS as "leaders" and were informed about the product.[15] Therefore leaders, or people close to them in the network, are more likely to have heard of the microfinance opportunity and have taken it up (Banerjee et al., 2013). Motivated by this, we estimate the random forest model based on household demographics and network characteristics from the microfinance villages on a training sample of 7199 households and then validate the method on a testing sample of 2399 households, with the training/testing splits in line with the literature (Stone, 1974; Breiman and Spector, 1992; Xu and Goodacre, 2018). The features used for classification are: (1) a dummy for whether the household has a female of eligible age (between 18 and 57), which BSS set as a requirement to be able to participate in microfinance; (2) a dummy for BSS leader households which are households that were specifically informed about the product when they entered the village; (3) the average closeness (mean of inverse of network distance) to leaders, which is relevant, because as in Banerjee et al. (2013), those who are closer to leaders should be more likely to hear of microfinance; (4) the average closeness (mean of inverse distance) to same-caste leaders, because interactions within caste are more likely and therefore should influence the likelihood of being informed; and (5) the share of same-caste leaders in the village. A description of the algorithm appears below (Algorithm 1). The details of the estimation algorithm, implemented choices, and quality are presented in Appendix B.

Turning to the Hyderabad setting, the strategy is similar, though Spandana had a more multifaceted approach for selecting borrowers. Thus, we consider 19 predictors of a household's take-up of Spandana, including demographic characteristics of the household (such as characteristics of the household head and his spouse, the number of women and children in the house, whether the household owns a business) as well as demographic data for the village (such as literacy rate, village population, total number of businesses in the village). We again use random forest, training a model on 2520 households and then validating the model on a testing sample of 1080 households.

We next apply the classifier to both microfinance and non-microfinance villages (or neighborhoods) to classify each household as H or L (high or low likelihood of joining microfinance).

An advantage of using random forests is that they naturally allow for non-linearities and potentially complex interactions between characteristics that could drive microfinance take-up.[16] A related advantage of random forest comes from its value in identification. Because

[15]The BSS definition of leader was defined by occupation (e.g., teachers, self-help group leaders, shopkeepers), so we can identify them similarly in MF and non-MF villages.

[16]Alternatives such as logistic regressions would not typically be able to handle such interactions and non-linearities without introducing very high dimensional interaction terms.

Algorithm 1: Random Forest

Data Input:

- N = Set of respondents from all villages,
- N_{mf} = Set of respondents from microfinance villages,
- Y_i = Loan take-up binary outcome for each $i \in N_{mf}$,
- X_i = Set of predictor variables for each $i \in N_{mf}$.

Parameters:

- T = Set of trees to grow,
- p = Total number of predictors,
- m = Number of predictors selected at each split,
- c = Cut-off: vector of length 2 (the winning class for an observation is the one with the maximum ratio of proportion of votes to cut-off),
- t = Fraction of sample to be used as training dataset.

(1) Randomly select (with replacement) training data S and testing data S' from N_{mf}. The size of S will be $t \cdot n(N_{mf})$ and the size of S' will be $(1-t) \cdot n(N_{mf})$.

(2) For each tree $t \in T$,
   - Randomly select (without replacement) a sample of size $n(S)$ from S.
   - At each node n of the tree t, randomly select (with replacement) a set of predictors of size m from p.
   - At each node, construct a split using Gini's Diversity Index.
   - For every tree t, each $i \in N_{mf}$ will be assigned a classification $\hat{Y}_{it} \in \{0,1\}$.

(3) After classifying each $i \in N_{mf}$, for each tree t, the final classification is computed as follows,

$$\hat{Y}_i = 1\left\{\frac{1}{n(S)}\sum_{t=1}^{n(T)} \hat{Y}_{it} > c[2]\right\}$$

and therefore $\theta_i = \hat{Y}_i \cdot H + (1-\hat{Y}_i) \cdot L$.

random forests allow for classification via a complicated non-linear function of the network and relation to leadership positions, in the Karnataka setting, where we have baseline network data, we can control smoothly for network position and network position interacted with post. Therefore, unobservables correlated smoothly with network parameters are unlikely to drive the Karnataka results.

Random forest classification does have a few downsides. In our case, the main one is that if the true underlying data-generating process has log-odds that are linear in parameters, then the random forest may overfit. Therefore, for robustness, we also present our main results in Appendix J using logistic regression to classify households into H and L types for both Karnataka and Hyderabad.[17] All of the results are replicated in both samples. In Appendix Section B.5, we show that random forest performs comparably to logit in Karnataka, while random forest outperforms logit on all metrics in Hyderabad.

[17]In Table 1, we report evidence of baseline imbalance in the random forest classifier in Karnataka. We detect no such imbalance using the random forest classifier in Hyderabad or the logit classifier in either sample.

Table 2 presents some summary statistics from the classification exercise. In Panels A and B we look at Karnataka data. There are notable differences between H and L households. Although none of these features were used in the estimation, we find that H households are much more likely to be SC/ST, have smaller houses in terms of room count, are much less likely to have a latrine in the household, and are much less likely to have an RCC (reinforced concrete cement) roof, all of which suggests that they tend to be poorer. Finally, we see that H households have somewhat larger network degrees than L households, and the composition exhibits homophily: H types have a lower number of links to L types and a higher number of links to H types. Finally H households are more eigenvector central in the network, which is not surprising given that they were selected in part based on being closer to BSS leaders, who themselves tend to be more central. In Section 5.1 we show that indeed H types borrow considerably more than L types in microfinance villages. H types borrow Rs. 1,785 more than L types in Karnataka ($p < 0.001$), indicating that the classification performs well.

Panels C and D turn to the Hyderabad data and look only at the non-microfinance villages. In Panel C, in contrast to Karnataka, we do not find a pattern of significant differences between H and L households in their demographic characteristics. Turning to network characteristics, in Panel D we see, like in Karnataka, that H types have fewer links to L types, more links to H types, and are more central. Again, in Section 5.1 we show that, one year after microfinance entered the treated neighborhoods, H types had considerably more microcredit than L types in early microfinance neighborhoods (Rs. 8,773, $p < 0.001$).

3. Changes in Networks

How does exposure to microfinance change networks? We begin with a discussion of how the overall structure of social networks are affected and then discuss the effects on different types of bilateral links as well as triads.

3.1. **Effect on the total number of links.** To look at how introducing microfinance affects the overall structure of village social networks, in the Karnataka data, where we have network panel data but no randomization, we use a difference-in-differences framework:

$$y(\mathbf{g}_{vt}) = \alpha + \beta \text{Microfinance}_v \times \text{Post}_t + \gamma \text{Microfinance}_v + \eta \text{Post}_t + \delta' X_v + \epsilon_{vt},$$

where $y(\cdot)$ computes the density of the network $\mathbf{g}_{vt}$ for village v in period t, the average closeness (the mean of the inverse distance between all pairs), or clustering. The density is the percentage of links a random household has to all other households in the village, so it measures how well-connected the village is on average.[18] The distance in the network is the (minimum) number of steps through the network it takes to get from one household to another. In models where favors, transactions, or information travel through the network, higher distance or lower closeness (the

[18]Note that density is directly related to average degree—it is proportional to average degree scaled by $n - 1$.

inverse) means that the movement of such phenomena through the network is slower. Finally, clustering is the share of a household's connections that are themselves connected. Economic models of network formation identify clustering as an important feature to sustain cooperation. X_v is a vector of control variables, which varies according to the specification as discussed below.

Table 3, panel A presents the results for Karnataka. Columns 1-3 present results for network density, columns 4-6 for clustering, and 7-9 for closeness. The first column in each category (columns 1, 4, 7) presents a simple difference in differences specification without controls. The second column in each specification (2, 5, 8) adds to that a vector of baseline controls interacted with Post_t. These controls include share of upper-caste households, number of households in the village, network density, share of households in self-help groups, share Hindu, share with a latrine in the house, share that own the household, share that have electricity and share that are leaders. We add these because differences in the size of the village, its caste composition, or the wealth distribution could potentially have effects on the evolution of networks even without introduction of microfinance. While the entry of BSS does not seem to correlate with much of anything beyond village size, we include these controls to ensure that they do not drive the results. Finally, the third column in each specification (3, 6, 9) includes village fixed effects as well as controls for the baseline value of the outcome variable interacted with Post, to allow for differential time trends by baseline network feature. Because we only have 150 observations but many controls (up to 18 controls and their interactions before adding the fixed effects), we use the double post-LASSO (DPL) procedure (Belloni and Chernozhukov, 2009; Belloni, Chernozhukov, and Hansen, 2014a,b) to select the controls.[19]

We find that exposure to microfinance leads to a drop in network density by about 1.2-1.3pp relative to a mean of 11.4% in non-microfinance villages in Wave 1 (columns 1-3, $p = 0.077$ in column 3 for example). This is an 11% drop in density. We do not find any detectable effect of microfinance on clustering in the villages. This is true irrespective of whether controls are used. Without controls we find a significant reduction in the average closeness (column 7, $p = 0.02$), corresponding to a 0.53 standard deviation effect. However, this loses significance in columns 8 and 9 with the inclusion of controls ($p = 0.19$, $p = 0.21$, respectively).

Panel B turns to the Hyderabad data, which uses an endline cross-sectional dataset rather than a panel, but takes advantage of the random selection of neighborhoods to treatment. There, we run the following specification.

$$y(\mathbf{g}_{vt}) = \alpha + \beta \text{Microfinance}_v + \delta' X_v + \epsilon_{vt}$$

where the vector of controls X_v are demographic characteristics of the household and the village, the same controls used for classification of H and L. We again use DPL to select the control

[19]Because the double post-LASSO procedure does not select all of the village fixed effects, we can include the fixed effects and an indicator for microfinance in the same regression.

variables. We find that there is a 21% decline in density ($p = 0.086$ without DPL in column 1 and $p = 0.062$ with DPL in column 2). We do not find meaningfully significant results on clustering or closeness (columns 3-6). In both settings, therefore, we find a reduction in the overall density of the network in response to microfinance exposure.

3.2. **How are links affected by microfinance?** In this subsection, we explore how microfinance exposure affects the formation of links across types of households.

Before turning to results using the predicted Hs and L types, we can first conduct a simple back-of-the envelope exercise to gauge whether it is plausible that the full drop in density could have come only from drops in links involving microfinance takers (i.e., Hs). Note that in Karnataka only 18.5% of households in treatment villages took up microfinance (Banerjee et al., 2013), implying that the majority of links in the network connect pairs of non-takers. If the drop in density measured in Table 3 only affected links including at least one microfinance taker, then our village-level estimates would correspond to a 47% decline in these types of links.[20] This decline — corresponding to a loss of nearly half of baseline links — is implausibly large, suggesting that microfinance *must* have also impacted relationships between non-takers.[21]

To make more progress, we make use of the classification of households into H and L types, introduced in Section 2.4. Bilateral links can be of three types: HH, LH, and LL. Let $g_{ij,v,t}$ be an indicator for whether a link is present between households i and j in village v in wave t. Letting LH_{ij} be an indicator for pair consisting of one low type and one high type, and analogously for HH_{ij} etc., the regressions we run take the form

$$\begin{aligned} g_{ij,v,2} = \alpha &+ \beta MF_v + \beta_{LH} MF_v \times LH_{ij,v} + \beta_{HH} MF_v \times HH_{ij,v} \\ &+ \gamma_{LH} LH_{ij,v} + \gamma_{HH} HH_{ij,v} + \delta' X_{ij,v} + \epsilon_{ij,v,2}, \end{aligned}$$

where $X_{ij,v}$ includes a vector of flexible controls (a polynomial) for centrality of both nodes, demographic variables (caste and wealth proxies including number of rooms, number of beds, electrification, latrine presence and roofing material), all variables that are used in the random forest classification; and interactions of all of these variables with the microfinance indicator (the control variables finally included are chosen by DPL).

The idea behind identification is that the classification type, H or L, is a complex, non-linear function with many interaction terms of a subset of the features described above. As such,

[20]In the absence of microfinance, the wave 2 density in treatment villages would have been 0.08, measured as the wave 1 density in treatment villages plus the wave 2 effect estimated using the difference-in-difference specification. If we assume all potential links are equally likely, we can calculate the implied number of links involving microfinance takers for the average village (with 223 households) in the absence of microfinance, which is 678. The density reduction observed in microfinance villages corresponds to an overall reduction of 317 total links, which would represent 47% of the total number of counterfactual links involving microfinance takers.

[21]Note that the relative decline in density is even larger in Hyderabad for a smaller treatment effect on take-up, suggesting an even greater decline in links involving takers would be required to explain the entire observed reduction in links.

we can still smoothly control for the features and allow the control to vary by whether the village is exposed to microfinance or not. This allows us to control for potentially differential effect of microfinance exposure on households that are demographically distinct and located differently in the network, under the maintained assumption that these effects can be captured by linear uninteracted terms. The coefficients of interest, β_{LH} and β_{HH}, capture whether being in a microfinance village differentially affects the evolution of a link among types classified as LH and HH, relative to those classified as LL, conditional on all the characteristics above and their interaction with MF. We also present regressions without any controls whatsoever to demonstrate that the results are robust to the presence or absence of these detailed controls. Altonji, Elder, and Taber (2005) show that if results do not change as more and more controls are introduced, then this provides support for the view that unobservables are not spuriously driving the results.

We run these regressions for two samples: the set of ij such that $g_{ij,v,1} = 1$ (in this case we ask whether pre-existing links break) and the set of ij such that $g_{ij,v,1} = 0$ (so the link doesn't exist in the first period), in which case we ask about the probability of a new link forming in Wave 2.

Table 4 presents the link-level results for any type of relationship in the Karnataka data. In columns 1-2 we focus on the set of links existing in Wave 1, and in columns 3-4 we focus on the set of unlinked nodes in Wave 1. Columns 1 and 3 include no controls whatsoever, and columns 2 and 4 introduce the set of control variables, and their interaction, with MF, selected by double-post LASSO. The key coefficients for testing the hypotheses are β, the coefficient on Microfinance, which captures the effect on the omitted category of links (LL), as well as β_{LH} and β_{HH}, which ask whether microfinance effects are different for these types of links, compared to LL. Column 1 shows that LL links break significantly more in MF villages relative to non-MF villages. Specifically, they are 5.8pp less likely to exist in Wave 2 ($p = 0.002$), relative to a base of 48.2% in non-MF villages. The reduction in LH links is very similar, while the HH links are less likely to disappear (and on average the HH links are not statistically more likely to break in MF villages than in non-MF villages, $p = 0.361$). The results are robust to the inclusion of control variables.

Columns 3 and 4 present similar results for link formation. LL links are 2.3pp less likely to form in microfinance villages on a base of 7.5% in non-microfinance villages ($p = 0.006$ without controls and $p = 0.008$ with controls). Again, the effect is comparable for LH links, i.e., β_{LH} is small and not significant (either with or without controls). The negative effect on link formation is somewhat less pronounced for HH links: β_{HH} is positive and significant with controls (but not significant without controls.

The differential changes in network structure in the microfinance villages shed light on network formation. The fact that the LH links break may reflect the fact that the Hs are no longer interested in maintaining their links with the Ls now that they have an alternative source of credit. The fact that LL links are equally likely to break, and fail to form, is more surprising, especially since the Ls should have a stronger incentive to hold on to their mutual links precisely because they no longer have access to the links with the Hs.[22]

We turn to the Hyderabad data in Table 5. In this case, while microfinance access is randomized, we have only cross-sectional information on networks so we cannot condition on pre-period link status. Therefore we run the regression in the sample of any possible link ij. The "microfinance" coefficient identifies the effect on LL links (the omitted category) and captures a net effect driven by both link formation and link destruction. Column 1 includes only the randomization strata as controls, while column 2 additionally allows for any of the household or village level controls used in the random forest classification to be included. In column 1, we find a 0.5 percentage point (on a base of 2.7 percent) decline in the probability that an LL link exists in microfinance neighborhoods relative to non-microfinance neighborhoods ($p = 0.035$). We cannot reject that the estimates for LH and HH are the same, but they are imprecisely estimated. The estimates are quite similar in column 2, after adding controls.

We next unpack these findings by distinguishing financial links (those that we anticipate would directly be affected by the credit injection) versus information links. Table 6, Panel A presents the results in the Karnataka data, where columns 1 and 2 consider the evolution of financial links, while columns 3 and 4 consider informational links. Columns 1 and 3 restrict to links of each type that existed in the Wave 1 data, while columns 2 and 4 restrict to pairs of individuals that were not linked in Wave 1. The patterns are strikingly similar across financial and information links — both types of LL links are significantly more likely to break (cols 1 and 3) and significantly less likely to form (cols 2 and 4), which is evidence of multiplexing. In fact, for information links, we find that the disappearance of HH links is significantly smaller than that of LL links (column 3, $p = 0.063$).

In the Hyderabad data, recall that the link-level information analyzed in Table 5 is constructed using ARD, which defines a link as a relationship along any dimension (information, advice, or financial), so we cannot run an exactly parallel specification. However, we did collect supplemental, node-level information on relationships, by type, that allows us to explore whether microfinance affects financial and non-financial (i.e., advice and informational) links differentially. Panel B of Table 6 presents household-level regressions where the outcome variable is the number of financial or non-financial links, and the regressors are MF and $MF \times H$ (with and without control variables). The main effect of MF identifies the effect of microfinance exposure for L households. It is negative and highly significant on the number of financial links ($p = 0.01$ without controls in col 1 and $p = 0.006$ with controls in col 2). For the number of

[22]In Online Appendix K, we show that these impacts are robust to differential trends by village size interacted with link type. We also show a specification interacting treatment with each of the controls.

non-financial links, the effects are negative and non-negligible in magnitude, but not quite significant at conventional levels ($p = 0.101$ without controls in col 3 and $p = 0.169$ with controls in col 4). Still, there is clear evidence of multiplexing effects for the Hs, with the changes in links without and with controls significant ($p = 0.045$, $p = 0.013$, respectively).

3.3. **Group Relationships.** In the link-level analysis we show that LL links — relationships between two individuals who are likely to experience very limited, if any, *direct* impacts from microfinance — are at least as likely to be affected as relationships involving Hs. One natural place to look first to try to understand this result is local payoff externalities: does the decline in LL links stem from these households' mutual links to other H households who join microcredit?

Bloch, Genicot, and Ray (2008); Ambrus, Mobius, and Szeidl (2014); and Jackson, Rodriguez-Barraquer, and Tan (2012) all propose models where contract enforcement requires groups of nodes rather than simple pairs. In Jackson et al. (2012), for example, two households seeking to exchange favors may not have enough bilateral interaction to be able to sustain cooperation in isolation. However, if they both have relationships with some other households in common, then the relationships can all "support" each other and provide incentives to cooperate: if someone fails to cooperate with one of their friends, then beyond losing that relationship, they can also lose relationships with all the other friends they had in common.

Our network data exhibits such groupings, with the likelihood of a group of nodes being collectively linked far exceeding the likelihood to be expected if decisions were made independently (see Online Appendix H.2). These interdependencies in link formation could in principle explain the impact of microfinance on LL links. If there are payoff externalities, L types might value an LL link more when there is a third node involved. The introduction of microcredit could destabilize these structures. In groups that are composed of both L and H types, it could be the case that microfinance directly causes LH links to break, which in turn spills over to adjacent LL links in the same group. In this world, groups comprising only Ls should experience minimal impacts.

The direct testable implication is that, if we focus on triangles that existed in Wave 1, we should see a larger decline in triangles involving at least one H than in LLL triangles. We use the following specification to test this hypothesis:

$$\begin{aligned} y_{ijk,v,2} = {} & \alpha + \beta MF_v + \beta_{LHH} MF_v \times LHH_{ijk,v} + \beta_{LLH} MF_v \times LLH_{ijk,v} + \beta_{HHH} MF_v \times HHH_{ijk,v} \\ & + \gamma_{LHH} LHH_{ijk,v} + \gamma_{LLH} LLH_{ijk,v} + \gamma_{HHH} HHH_{ijk,v} + \delta' X_{ijk,v} + \epsilon_{ijk,v,2}, \end{aligned}$$

where $y_{ijk,v,2}$ is either a dummy for whether the triangle ijk exists in Wave 2 ($g_{ij,v,2} g_{jk,v,2} g_{ik,v,2} = 1$) in some specifications, or whether any link in the former triangle exists in Wave 2 ($g_{ij,v,2} + g_{jk,v,2} + g_{ik,v,2} > 0$) in other specifications. The vector $X_{ijk,v}$ includes flexible controls for centralities of households, the demographic characteristics previously described for all households, all classification variables used in the random forest model and the interactions of all of these variables with microfinance. As before, we present regressions with and without controls.

Table 7 presents the results in the Karnataka data. In column 1, we find that LLL triads are 7.8pp more likely to dissolve in microfinance relative to non-microfinance villages ($p = 0.008$). The positive coefficients on the terms β_{LLH}, β_{LHH} and β_{HHH} indicate that these triads are less likely to dissolve, significantly so in the case of LHH and HHH. The results are similar with control variables (col 2), though the differences across types of triples are less precise. Similarly, in column 3, we see that, among formerly-linked triangles, we are more likely to see that none of the links survive for LLL triangles in MF villages (-8.5 pp, $p \leq 0.001$), and that this is significantly less likely to occur for LLH, LHH, and HHH triangles.

Table 8 presents the Hyderabad results and measures whether microfinance affects the likelihood that a given set of three households are all linked (recall that we do not have baseline data, so we cannot condition on pre-existence). Because the likelihood of any potential triangle being fully linked is low (approximately 0.01%), we scale all regressors by 1,000 for readability.[23] Although the results are noisier than in Karnataka, we find once again that LLL triangles are negatively affected by microfinance: in column 1, we are 50% less likely to see any LLL triangle in MF neighborhoods ($p = 0.067$). The effect is statistically indistinguishable from that on LLL triads for LHH and LLH triads. However, unlike in Karnataka, we do find some evidence that the HHH triangles are more likely to drop than the LLL triangles ($p = 0.067$ without controls and $p = 0.139$ with controls).

In summary, we find that LLL triangles decline at a rate similar to triangles containing at least one H node, despite the fact that in an LLL triangle, none of the nodes are *directly* affected by microfinance. This suggests that simple models of local externalities cannot explain our results. The next section proposes an alternative model that can rationalize these findings.

3.4. **Classification and ARD Robustness.** In Appendix F, we present a discussion of the robustness of the ARD classification. Specifically, we conduct two exercises. First, we show that our results are not sensitive to the specific training data used to construct our L, H classifier. Second, we show that our Hyderabad results are robust to incorporating estimation error from the ARD procedure into our regression analysis.

4. A Model

In this section we present a new model of network formation, in which the externalities are global rather than local. These arise because of the nature of the search process by which people make and maintain friendships (see the discussion in Appendix D). This is in contrast with local externalities where the value of a link depends only on the other connections of the people being linked. In reality, both types of externalities, local and global, are surely at play. Local factors—being introduced to a job by friends of friends, etc.—are important for many aspects of life. At the same time, more undirected socializing effort—going to the town

[23]This sparsity of groups of triangles also implies that a pooled cross-sectional analysis will largely reflect new link creation rather than existing link maintenance.

square or local lea shop to chat and catch up on news—is also clearly relevant to the way that information spreads and that friendships are formed and maintained. We focus on this latter possibility which has so far not been emphasized in the literature.

We present the model for links and then describe how it can be extended to cover triads. As the model may be useful beyond the current setting of microfinance, we describe it in a general form and then specialize to the two-type (H, L) microfinance case.

4.1. **Types and Utilities.** There are n individuals, indexed by $i, j... \in \{1, \ldots, n\}$. Agent i has a type θ_i from a type set Θ. Let $v_{\theta\theta'}$ denote the base benefit that an agent of type θ gets from a relationship with an agent of type θ'. For example, in our context, this can come from borrowing and lending activities, as we discuss in more detail below.

The realized utility from a relationship also involves an idiosyncratic noise term ε_{ij} that i gets from being friends with j. This could be personality compatibility or some other benefits. Thus, an agent i gets a value $v_{\theta_i\theta_j} + \varepsilon_{ij}$ from a connection with j, where ε_{ij} is distributed according to an atomless distribution F.

A useful expression is

$$\mathrm{E}^+[v] = \mathrm{E}\left[v + \varepsilon_{ij} \left| \varepsilon_{ij} > -v\right.\right] = v + \frac{\int_{-v}^{\infty} \varepsilon_{ij} dF}{\int_{-v}^{\infty} dF},$$

which denotes the expectation of $v + \varepsilon_{ij}$ conditional the value of $v + \varepsilon_{ij}$ being positive. This is the expected utility that an agent gets from a relationship with base value v, conditional upon being willing to form the friendship.

An agent of type θ then has an expected utility from $d_{\theta\theta'}$ friends of type θ' of

$$\sum_{\theta' \in \Theta} d_{\theta\theta'} \mathrm{E}^+[v_{\theta\theta'}]. \tag{4.1}$$

4.2. **Efforts and Link Formation.** Each agent chooses an effort $e_i \in [0, 1]$, which represents the amount of time they spend socializing to form and maintain links. In the case of the villagers, this could be time spent in the town square or tea shop, where they meet with other villagers.[24] As will become evident, our model is meant to capture both link formation and link maintenance.

Two agents i and j who have chosen efforts e_i and e_j have probability proportional to $e_i e_j$ of meeting. The model therefore rules out "directed search" since the probability of meeting is independent of the agent's type, conditional on their effort. Time goes in periods $t \in \{0, 1, 2, \ldots\}$.

Let $g^t \in \{0, 1\}^{n \times n}$ be the adjacency matrix representing network at time t. If $g_{ij}^{t-1} = 1$, then those agents were connected in the last period and they keep their friendship if they meet with each other during time t — keeping the relationship requires seeing each other. Therefore

[24]This is a useful and conventional modeling device. See Currarini, Jackson, and Pin (2009, 2010); Cabrales, Calvó-Armengol, and Zenou (2011); Canen, Jackson, and Trebbi (2017) for other models where socialization takes effort and there is random meeting.

agents i and j keep their friendship with probability $e_i e_j$ and lose it with probability $1 - e_i e_j$. If $g_{ij}^{t-1} = 0$, then agents i, j were not connected in the last period and form a friendship with probability

$$e_i e_j \left(1 - F(-v_{\theta_i \theta_j})\right) \left(1 - F(-v_{\theta_j \theta_i})\right).$$

This is the probability that they meet *and* they both find the friendship of positive value – a friendship requires mutual consent in our model.

Thus, the efforts of agents do two things: they maintain old relationships by continuing an interaction but also allow them to meet new people.

4.3. **Steady-State Equilibrium.** A *steady-state equilibrium* is a list of efforts $(e_\theta)_{\theta \in \Theta}$, and a corresponding set of expected degree levels $(d_{\theta\theta'})_{\theta\theta' \in \Theta^2}$ such that e_θ maximizes each agent's expected utility, and the expected degree levels are in steady state as generated by the efforts.[25] We prove in the Appendix A that in all equilibria all agents of the same type choose the same action and that the equilibrium is unique, provided that costs of effort are not too small.

The requirement that degrees are in steady state and generated by the efforts is represented as follows. Let $n_{\theta\theta'}$ denote the number of agents of type θ' with whom an agent of type θ could potentially form friendships. If $\theta' \neq \theta$, then this is generally the number of agents of type θ',[26] while if it is of type θ, then it is less by one to account for the agent herself.

Out of those agents only an expected fraction of $(1 - F(-v_{\theta\theta'}))(1 - F(-v_{\theta'\theta}))$ will be friends with an agent of type θ, given the mutual consent requirement. Thus, let

$$m_{\theta\theta'} = n_{\theta\theta'} (1 - F(-v_{\theta\theta'})) (1 - F(-v_{\theta'\theta})).$$

This is the expected size of the pool of agents of type θ' with which an agent of type θ could be friends over time.

Degree at the end of a period is the maintained relationships plus the new ones formed:

$$d_{\theta\theta'} = e_\theta e_{\theta'} d_{\theta\theta'} + (m_{\theta\theta'} - d_{\theta\theta'}) e_{\theta'} e_\theta,$$

which simplifies to

$$d_{\theta\theta'} = m_{\theta\theta'} e_\theta e_{\theta'} = n_{\theta\theta'} (1 - F(-v_{\theta\theta'})) (1 - F(-v_{\theta'\theta})) e_\theta e_{\theta'}.$$

Thus, in steady state, degree is proportional to the number of available agents of the other type, weighted by the probability that there is a mutual compatibility and by the socializing efforts.

[25]We solve the model in terms of steady-state and expected values, but it will be clear from the analysis that one can also do this in terms of realized values. The equilibrium will still be unique for sufficiently high costs of effort, complementarities still apply in the same manner, and the equilibria have the same comparative statics. The complication is that strategies need to be specified as a function of more than just type, as the realized noise terms then matter. Since the noise terms add no insight, we work with this more transparent version.

[26]It could also incorporate some other taboos or restrictions, for instance if some types simply are not permitted to form relationships, which would be captured by the vs.

The expected utility of an agent involves the benefits from relationships, the costs of socialization, $-\frac{1}{2}c_\theta e_\theta^2$, as well as a benefit just from from socializing, $u_\theta e_\theta$. An agent may get some value from going to the town square or getting tea, etc., independently of who else is there.

This leads to a utility of

$$V_\theta(e_\theta) = \underbrace{u_\theta e_\theta - \frac{1}{2}c_\theta e_\theta^2}_{\text{base socializing benefit and cost of effort}} + \underbrace{\sum_{\theta' \in \Theta} \mathrm{E}^+[v_{\theta\theta'}] d_{\theta\theta'} e_{\theta'} e_\theta}_{\text{expected maintenance of existing friendships by effort}}$$

$$+ \underbrace{\sum_{\theta' \in \Theta} \mathrm{E}^+[v_{\theta\theta'}] \left(m_{\theta\theta'} - d_{\theta\theta'}\right) e_{\theta'} e_\theta}_{\text{expected new friendships from effort}}$$

Using the expressions for $m_{\theta\theta'}$ and $d_{\theta\theta'}$, this is

$$V_\theta(e_\theta) = u_\theta e_\theta - \frac{1}{2}c_\theta e_\theta^2 + \sum_{\theta' \in \Theta} \mathrm{E}^+[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right) e_{\theta'} e_\theta.$$

If we take $u_\theta \geq 0, c_\theta > 0$ for all θ and $\mathrm{E}^+[v_{\theta\theta'}] \geq 0$ for all θ, θ', then an equilibrium requires that:[27]

$$e_\theta = \min\left\{1,\ \frac{1}{c_\theta}\left(u_\theta + \sum_{\theta' \in \Theta} \mathrm{E}^+[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right) e_{\theta'}\right)\right\}.$$

4.4. **Equilibrium Existence and Some Comparative Statics.** This is a game of strategic complements, and for such games equilibria exist and form a complete lattice.[28] If $u_\theta = 0$ for all θ, then there exists a corner equilibrium in which all agents exert 0 effort. To examine the more interesting case, we presume that $u_\theta > 0$ for all agents, so that agents gain some utility from socializing regardless of the connections they form from it. In this case, for high enough costs of socialization there exists a unique equilibrium which has the property that there are spillovers from a change in the preferences of any type on the effort choices of all other types. The following result is proven in Appendix A.

Proposition **1.** *Let $u_\theta > 0, c_\theta > 0$ for all θ. For sufficiently large $c_\theta > 0$'s, there is a unique equilibrium. This equilibrium is stable*[29] *and interior ($0 < e_\theta < 1$ for all θ), and agents of the same type take the same efforts. In addition, if $\mathrm{E}^+[v_{\theta\theta'}] > 0, n_{\theta\theta'} > 0$ for each θ, θ',*[30] *and $v_{\theta\theta'}$*

[27]These come from the first order conditions, capped by the bound on efforts. Second order conditions are $-c_\theta$ and so are negative. Thus, these conditions are also sufficient.

[28]For instance, see Van Zandt and Vives (2007).

[29]Slight perturbations of efforts lead to best reply dynamics that converge back to the equilibrium.

[30]All that is needed for this result is that this holds for a cycle of θ and θ' that include all types. Note also that $\mathrm{E}^+[v_{\theta\theta'}] > 0$ does not require that all people form links, just that there is a non-zero probability that any two types could find a high enough noise term to form a friendship.

is decreased for some $\theta\theta'$ (holding all other parameters constant), then $e_{\theta''}$ decreases for all θ'', and $d_{\theta''\theta'''}$ decreases for all $\theta''\theta'''$.

The characterization of equilibrium is as follows. Let u be the $|\Theta|$-dimensional vector with entries $\frac{1}{c_\theta}u_\theta$ and E be the $|\Theta| \times |\Theta|$ matrix with θ, θ' entries

$$\frac{1}{c_\theta}\mathrm{E}^+[v_{\theta\theta'}]n_{\theta\theta'}\left(1 - F(-v_{\theta\theta'})\right)\left(1 - F(-v_{\theta'\theta})\right).$$

Then the unique equilibrium is given by

$$e = (I - E)^{-1}u,$$

which we show is well-defined for large enough costs in Appendix A.

A major implication of the proposition is that a reduction in the returns from any type of relationship decreases *all* efforts and degrees. The decrease in value $v_{\theta\theta'}$ for some $\theta\theta'$ directly affects their efforts. Then, given the strict strategic complementarities, there is then a decrease in other efforts; and the feedback can lead to a substantial drop in all efforts.

Note that the relative drops in efforts and degrees predicted in Proposition 1 are not necessarily ordered across groups: degree can fall most among groups of nodes that experience no direct decline in link valuation (e.g., LL links in the case of microfinance). The intuition is that if marginal benefits to Ls from connecting to Hs are particularly high, then when Hs drop effort, payoffs from effort for Ls can drop even more than for Hs, leading to an even larger effect on Ls. See Online Appendix M for a simulation demonstrating this phenomenon.

4.5. **Externalities in Network Formation.** Even though our model does not include direct externalities in payoffs between links, the network-formation process still exhibits significant external effects since agents' decisions to form links (their effort levels) affect others' potential payoffs and their network formation decisions (e.g., agent 3 putting in less effort lowers the efforts of both agents 1 and 2 and thus the chance that 1 and 2 are linked, even though the potential $1-2$ link has nothing to do with agent 3).

This makes a point beyond the current setting: network formation can be inefficient not simply because of direct externalities in relationships, which is how it is usually modeled,[31] but also because meeting people requires effort, and this naturally generates strategic complementarities and substantial externalities.

4.6. **Specializing to Microfinance.** We next present a specialized case of the model to analyze how microfinance changes incentives for socialization. We first present the steady-state conditions for the case of two types, H and L. We then discuss how the introduction of microfinance maps to the model parameters, specifically the $v_{\theta,\theta'}$s.

[31]This class of models can incorporate, *inter alia* risk sharing, information sharing, and network support. For references see Jackson (2003, 2008).

4.6.1. *Two Types.* We now specialize the model to the case of two types: $\Theta = \{H, L\}$. Let λ be the share of H types in the population. In this case, a steady-state is a solution to the equations:

$$\begin{aligned} c_H e_H^\star = u_H &+ \mathrm{E}^+[v_{HH}]\,(\lambda n - 1)\,(1 - F(-v_{HH}))^2\, e_H^\star \\ &+ \mathrm{E}^+[v_{HL}]\,(1-\lambda)\, n\,(1 - F(-v_{HL}))\,(1 - F(-v_{LH}))\, e_L^\star, \end{aligned}$$

$$\begin{aligned} c_L e_L^\star = u_L &+ \mathrm{E}^+[v_{LL}]\,((1-\lambda)\, n - 1)\,(1 - F(-v_{LL}))^2\, e_L^\star \\ &+ \mathrm{E}^+[v_{LH}]\lambda n\,(1 - F(-v_{LH}))\,(1 - F(-v_{HL}))\, e_H^\star, \end{aligned}$$

$$d_{HL} = ((1-\lambda)\, n)\, e_H^\star e_L^\star\,(1 - F(-v_{HL}))\,(1 - F(-v_{LH}))\,,$$

$$d_{LH} = d_{HL}\frac{\lambda}{1-\lambda},$$

$$d_{HH} = (\lambda n - 1)\,(e_H^\star)^2\,(1 - F(-v_{HH}))^2\,,$$

$$d_{LL} = ((1-\lambda)n - 1)\,(e_L^\star)^2\,(1 - F(-v_{LL}))^2\,.$$

The equilibrium vector of efforts ($e_\theta^\star$s) and network structure (vector of $d_{\theta\theta'}$s) are determined by this system.

4.6.2. *Application to Microfinance: An Example.* How does the entry of microfinance affect these parameters? Here, we present a rationalization for the payoffs based on informal borrowing and lending. In particular, let the values $v_{\theta\theta'}$ be interpreted as "financial" payoffs from borrowing and lending. This tells us how $v_{HH}, v_{HL}, v_{LH}, v_{LL}$ change in response to Hs getting microcredit.

Lending produces a net return of r, which represents the effective expected interest rate from informal loans net of the opportunity cost of funds. Borrowing leads to a return net of repayment of b, which represents the difference between the interest rate charged by a network "friend" and the opportunity cost of foregoing that loan (e.g., losing the money or borrowing at some higher rate from a money lender, etc.). Generally, we expect $b > 0$ and $b > r$,[32] as otherwise such relationships make little sense. Whether r is positive or negative is not obvious, since there are clearly social expectations to help out friends in need (which could make r negative), and may depend on context.

A household can be in one of three states of the world: they have money to lend, they need to borrow, or neither. An H household has a probability α_H of having money to lend and a

[32]The limited evidence we have on peer-to-peer lending suggests that markups on loans to friends tend to be small, potentially even negative. b by contrast ought to be substantial and positive.

probability $\beta_H \leq 1-\alpha_H$ of needing to borrow, and with the remaining probability $1-\alpha_H-\beta_H$ neither occurs. There are similar probabilities α_L and β_L for the L types.

The base payoff to an agent of type $\theta \in \{H, L\}$ of being matched to agent of type $\theta' \in \{H, L\}$ is then

$$v_{\theta\theta'} = \alpha_\theta \beta_{\theta'} r + \beta_\theta \alpha_{\theta'} b.$$

As in the general model, we assume that expected utility is additive across all relationships (Equation 4.1) and that pairwise payoffs $v_{\theta\theta'}$ are independent of other friendships.[33]

The introduction of microfinance changes these parameters. There are several likely channels by which the introduction of microfinance affects the payoff parameters of H types from linking to others. Access to microcredit might impact both the demand and supply of informal loans by H types. If access to microcredit substitutes for informal loans, then we would expect β_H to decrease. If alternately, the weekly required repayments are burdensome to households, they may have to cut back on lending smaller sums to others in the village and may even start borrowing small amounts to repay the loans when cash is short, leading to a decline in α_H and perhaps an increase in β_H. In addition, if there are complementarities between formal and informal loans because receiving a MF loan allows the household to overcome a non-convexity[34], β_H could go up. In contrast, if re-lending of formal credit to network partners is common, a type H may have a probability $\alpha'_H \geq \alpha_H$ of being able to lend once she gets access to microfinance. Her probability of needing to borrow may also go down to $\beta'_H \leq \beta_H$ if microfinance loans are a substitutes for network credit. In any case, we maintain that the Ls' needs for borrowing and lending are unaltered by the introduction of microfinance. Let

$$\Delta\beta_H = \beta'_H - \beta_H \quad \text{and} \quad \Delta\alpha_H = \alpha'_H - \alpha_H$$

be the changes in the probabilities that the H types have borrowing and lending needs after microfinance. By our previous assumption, $\Delta v_{LL} = 0$.

Let $\Delta_{\theta\theta'}$ denote the resulting change in $v_{\theta\theta'}$. To get a feeling for how this depends on $\Delta\alpha_H$ and $\Delta\beta_H$, note that for small values of $\Delta\alpha_H$ and $\Delta\beta_H$, we get approximations

$$\Delta_{HH} = (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r+b) \qquad \Delta_{LL} = 0$$
$$\Delta_{HL} = \alpha_L \Delta\beta_H b + \beta_L \Delta\alpha_H r \qquad \Delta_{LH} = \alpha_L \Delta\beta_H r + \beta_L \Delta\alpha_H b.$$

As we describe above, the arrival of microfinance may impact the valuations through a range of mechanisms. Different mechanisms imply that $\Delta\beta_H$ and $\Delta\alpha_H$ could each be positive or negative, making it very hard to say anything general about the signs of these expressions. For example, consider the special case in which $\alpha_L = \alpha_H$, $\beta_L = \beta_H$ and $\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H = 0$. In this case $\Delta_{HH} = 0$. Meanwhile, as discussed above, we expect $b - r$ to be positive. Then

[33]We make this assumption to highlight our key mechanism of interest – that there can be spillovers to LL links that might be even bigger than spillovers to LH links despite not loading it into the payoff.
[34]See Banerjee et al. (2019a).

Δ_{HL} should be positive whereas Δ_{LH} should be negative as long as $\Delta\beta_H > 0$ and $\Delta\alpha_H < 0$. [35] Given that different valuations move in different directions, the prediction of how the different types should respond remains ambiguous.

However one obvious special case is when both α_H and β_H go down. In this case, as long as both b and r are positive, all of v_{HL}, v_{LH} and v_{HH} must go down. Thus, Proposition 1 applies, implying that $e^{\star}_H, e^{\star}_L, d_{HH}, d_{HL}, d_{LH}, d_{LL}$ should all fall. However, the relative declines in degree across groups can go in either direction.

In Online Appendix M, we use simulations to show that it is indeed possible that LL links may be most affected, given the decreased effort of Hs as well as the Hs' lower propensity to want to link with Ls given their decreased borrowing needs. Specifically, we consider the case where only v_{HL} declines when microfinance enters and look at the resulting declines in efforts as a function of the pre-microfinance levels of v_{HL} and v_{LH}. We demonstrate that in networks with homophily, where low types experience high value from linking to high types and high types experience substantially lower value from linking to low types, a decrease in H's valuation of links with Ls due to the entry of microfinance causes Ls to reduce socialization even more than Hs. This is because in equilibrium, mutual consent with the Hs is more likely to fail.

4.7. **Extensions of the model.** The model can be extended in several directions.

First, the model is solved in steady-state. Adding a population of unlinked (say "new-born") agents to the population of the unmatched is straightforward, as is having agents exit.

Second, note that it is plausible that when one aspect of a relationship becomes less important, there is some risk that the entire relationship breaks up, since there are costs to maintaining a relationship. By adding other types of links that are maintained and formed at the same time as financial links, the model can generate similar effects on other links as well. As we saw above in Table 6, when we look at advice-based links, the effects are more or less of the same magnitude in proportional terms and in the same direction as the financial links.

Third, we can extend the model to allow for triads and other dependencies, which we examine in detail.

4.7.1. *Adding More General Dependencies to Our Model.* We describe a variation on the subgraph formation model of Chandrasekhar and Jackson (2018).

Let G be some set of potential subgraphs on n nodes. For instance, instead of just a list of all possible links, it could also include triangles, or various other cliques, stars, and so forth.

We abuse notation and let $i \in g$ for some $g \in G$ denote that i is one of the nodes that has links in g. Let $v_i(g)$ denote the utility of i if g forms. The total utility that i obtains is the sum over all subgraphs that i is part of - so rather than just a network, the resulting object is a multigraph.

[35] See the calculations in Appendix O.

We let m_g denote a relative frequency adjustment for the type of subgraph in question, as some may be more or less likely to form as a function of the efforts.

The probability that some g forms if it is not present is then

$$m_g \times_{i \in g} e_i(1 - F(v_i(g)))$$

which is the product of the socialization efforts and the probability that each i involved in g finds it valuable to form g.

The probability that a subgraph is maintained if it is already present is[36]

$$\times_{i \in g} e_i.$$

Let $\mathrm{E}^+[v_i(g)]$ denote the expected utility that i gets from subgraph g conditional on finding it worthwhile to form, and $\mathcal{G}^t$ denote the set of subgraphs present at the beginning of time t. Then, the expected utility that i gets from effort e_i is

$$V_i(e_i) = u_{\theta_i} e_{\theta_i} - \frac{1}{2} c_{\theta_i} e_{\theta_i}^2 + \sum_{g \in \mathcal{G}^t : i \in g} \mathrm{E}^+[v_i(g)] \times_{j \in g} e_j$$
$$+ \sum_{g \notin \mathcal{G}^t : i \in g} \mathrm{E}^+[v_i(g)] m_g \times_{j \in g} e_j(1 - F(v_j(g))).$$

We say that a society is weakly connected in expectation if for all agents i and j, there exists a path between i and j if all $g \in G$ are formed. This is satisfied if at least one network that could conceivably form is path-connected, and is obviously satisfied if any two agents could be connected directly.

Proposition **2.** *Let $u_\theta > 0, c_\theta > 0$ for all θ, and the society be weakly connected in expectation. For sufficiently large $c_\theta > 0$'s, there is a unique equilibrium and it is stable and interior ($0 < e_i < 1$ for all θ). In addition, if $\mathrm{E}^+[v_i(g)] > 0, m_g > 0$ for each g, and $E^+[v_i(g)]$ is decreased for some g (holding all other parameters constant), then e_i decreases for all i, and $\mathrm{P}[g \in \mathcal{G}^t]$ decreases for all g.*[37]

From Proposition 2, it follows that, analogous to the dyads case above, all dyads and triads decrease in probability of forming if the value of at least one dyad or triad drops to some type. In particular, specializing the general model to the case of dyads and triads with types $\{H, L\}$, and corresponding values

$$v_{\theta\theta'}, v_{\theta\theta'\theta''}, (\theta, \theta, \theta') \in \{H, L\}^3,$$

we get the following corollary.

Corollary **1.** *Let $u_\theta > 0, c_\theta > 0$ for all θ. For sufficiently large $c_\theta > 0$'s, if $\mathrm{E}^+[v_{H,\theta'}]$ or $\mathrm{E}^+[v_{H,\theta',\theta''}]$ decreases for some $(\theta, \theta') \in \{H, L\}^2$, and no values increase, then the formation of dyads and triads of all types decrease.*

[36]One could adjust the relative impact of effort for maintaining a subgraph to be some other function than simply the product, depending on the context.

[37]The proof of Proposition 2 is easy to construct as an extension of that of Proposition 1, and thus omitted.

Again, in cases in which the value of connections from Ls to Hs, in either dyads or triads, is high enough, the drop in efforts by Hs can lead to a drop in L efforts that is large enough to lead to a drop in LLL triangles that is as large or larger than the drop in HHH triangles.

4.8. **Alternative Explanations.** In this section we try to address two issues. First, can we account for the facts without going to a model with undirected search while maintaining our assumptions about changes in payoffs? Second, are there alternative assumptions about changes in payoffs that can help us account for the facts in combination with a simpler model of network formation?

4.8.1. *Alternative models of network formation.* In Appendix D, we discuss four other models of network formation, variations of which are already in the literature. We argue that we need a model that goes beyond those models to account for the patterns in our data.

As such, our work contributes to the literature on network formation by introducing a model that combines features of different existing models, and showing why that combination of features is needed to match what we observe in the data.

Previous models of network formation that involve explicit choice by agents[38] have several flavors:

(i) models in which people have the opportunity to connect with whomever they want, subject to reciprocation (e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami (1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den Nouweland (2005); Herings, Mauleon, and Vannetelbosch (2009); Jackson, Rodriguez-Barraquer, and Tan (2012); Boucher (2015));
(ii) models in which there are exogenously random meetings and then, conditional upon meeting, people choose with whom to connect (e.g., Watts (2001); Jackson and Watts (2002); Christakis, Fowler, Imbens, and Kalyanaraman (2010a); König, Tessone, and Zenou (2014); Mele (2017));
(iii) models in which people put in some effort to socialize, which then results in some random meetings, but then relationships are formed as a result of those efforts without further choice (e.g., Currarini, Jackson, and Pin (2009, 2010); Cabrales, Calvó-Armengol, and Zenou (2011); Canen, Jackson, and Trebbi (2017)); and
(iv) models which emphasize local externalities such as payoffs from indirect connections with friends-of-friends (e.g., Bloch et al. (2008); Jackson (2008); Mele (2017); Badev (2013)).

First, the empirical patterns that we observe here require a model with some externalities in the efforts to search and meet, which are absent in (i) or (ii). In the basic models of this class, agents have full control over who they try to link with. This is the key difference with

[38]There is also a large literature of network formation that involves no strategic choice but just a stochastic model of network formation/evolution (e.g., see Jackson (2008) for some description and references). Those models are not equipped to match the data here.

a model of undirected search, and makes it difficult to explain why LL and LLL relationships drop in response to a decrease in Hs' willingness to link to Ls. Our model relies on the idea that individuals put effort into trying to meet but cannot choose who they are meeting. The reason Hs lower effort less than Ls is that they are (correctly) more optimistic about actually linking with those who they meet. If on the other hand, the Ls could costlessly meet with each other, or they meet people at random at no cost and can decide who to pair with, LL and LLL links should, if anything, go up.

Second, models in class (iii) allow for search efforts but then do not involve the choices of whom to connect to, as are present in (i) and (ii). This choice of whom to connect to is important in adjusting the incentives and the relative rates at which HL links form compared to HH or LL links, which is important for our results.[39] Thus, the model that we introduce is a hybrid of these three classes: effort is needed to meet others and affects the relative rates at which people are randomly met, but, conditional upon meeting, the two still have to decide to link.

Third, models in class (iv) involve encoding into the payoffs directly the value of maintaining friends-of-friends. While our model can easily accommodate this, as we describe in Section 4.7.1, this is not necessary to get the results we are interested in. Empirically, the result that LLL triangles are at least as likely to be affected as triangles involving Hs rules this out as a sole explanation. For example, one would need to construct a model where LLL triangles sustain even larger sub-groups of exchange (e.g., $HLLLH$ groups). Note, as one constructs externalities involving longer chains of nodes, it quickly becomes impossible to ever distinguish a global from a local externality. In the case of the Karnataka graphs, the average path length is under three, and the diameter is approximately five, so we believe it is reasonable to consider pairs and triples in our empirical analysis.

Fourth, our model has two other features that help us to match the data. One is that effort is not only needed to meet new people, but also to maintain existing relationships - as the patterns we observe in the data exhibit similarities both in terms of which relationships are retained and which new ones are formed. The second is that socializing affects the opportunities to form multiple types of relationships at the same time - and so relationships are naturally "multiplexed."

The combination of all four of these features – efforts to socialize with rates of meetings dependent on relative efforts, mutual choice required to form relationships conditional upon meeting, effort needed to maintain relationships, and multiple types of relationships formed at the same time – allows us to capture all of the nuances and rich patterns that we observe in the data. In Appendix D, we discuss why dropping any one of these features would fail to capture some aspects of the data.

[39]The models by Currarini, Jackson, and Pin (2009, 2010); Canen, Jackson, and Trebbi (2017) adjust the cross-type meeting rates either by a congestion meeting technology or a homophily parameter. Ours is derived from utility considerations, which helps understand why things change as microfinance is introduced.

4.8.2. *Alternative models of match value.* We have so far assumed that match value depends only on types and does not depend on the pattern of matching. It is possible, for example, that matches are substitutes, so that when many LH links break, the value of LL links may go up. This would predict an increase in LL links, which goes in the wrong direction. It is also possible, though perhaps less likely, that links are complements: perhaps when an L can no longer borrow from the Hs, she gives up the entire project and therefore also stops borrowing from other Ls. However in this case the LL links break because some LH links have disappeared and therefore the effect on LL links should be smaller than the effect on LH links in proportional terms.

A similar possibility is that an L might want to link with another L because that second L is in turn linked to an H, and this is valuable for another reason (e.g., risk sharing, information exchange, network support, etc.) But, as above, in this case the LL links are breaking because LH links have disappeared and so again, the effect on LL links should be smaller than the effect on LH links in proportional terms. See Appendix D.3.2 for more details.

Another possibility is that the reason LL links drop is that Ls recognize that even if they don't participate in microfinance, it is available to them. This is probably true for some of them, but because we use microfinance eligibility to determine who is an L, it is less true for them than for the H (who also don't all borrow). An H is therefore more likely to break their link with an L on these grounds than another L.

Yet another alternative is based on the idea that the very fact that Hs tend to socialize with Hs in microfinance meetings would provide a force unique to participants, hence Hs, to form new links. This might crowd out their other links, but that would predict that LH links should decline by more than LL links (which should not be affected). We do not find this. We further examine this alternative in Online Appendix I. We show that our main results hold even if we condition on all pairs where neither member joined microfinance (86% of baseline links). Even under this restriction, the H classification has content – Hs are more suitable for microfinance by construction, and therefore even non-borrowers have higher option value from future access to microfinance.[40] However, we acknowledge that these results are only suggestive given that actual microfinance take-up is endogenous.

Another possibility is a slight variant of our undirected search model where the H types simply do not have time to meet with the Ls anymore. Notice our general form of the model allows for this.

A final possibility is that the entry of microfinance leads to rapid economic growth in the village, so that both H and L types don't need to maintain informal relationships any more. This is not only inconsistent with the extensive literature on microfinance, which finds little impact of microfinance entry on average village or neighborhood level outcomes such as

[40]This option value may in turn reduce the value non-borrowing Hs receive from maintaining and forming network relationships.

consumption, investment or business profit (Angelucci, Karlan, and Zinman, 2015; Attanasio, Augsburg, De Haas, Fitzsimons, and Harmgart, 2015; Augsburg, De Haas, Harmgart, and Meghir, 2015; Banerjee, Karlan, and Zinman, 2015b; Banerjee, Duflo, Glennerster, and Kinnan, 2015a; Crépon, Devoto, Duflo, and Parienté, 2015; Tarozzi, Desai, and Johnson, 2015) (see Meager (2015) for a meta-analysis), but also with our findings in Section 5.2, below, that L households experience no change in income from greater community access to microfinance and a decreased ability to smooth consumption.

5. Informal Credit and Insurance

In this section we ask whether the changes in the networks documented so far are reflected in changes in households' economic outcomes.

5.1. Impact on Borrowing Patterns.

We begin by looking at how different types of borrowing respond to the arrival of microfinance. In both the Karnataka and Hyderabad data we have rich borrowing information, and we can measure the impacts of microfinance on several different types of household borrowing (e.g., microfinance, friend, self-help group member, family member, or money lender). If the loss in network links corresponds to a drop in informal financial transactions, then informal borrowing should respond in a manner similar to our network results above.

We have data on the amount borrowed by source for the entirety of our sample. We begin by regressing the amount borrowed on dummies for microfinance village, post, and household type in the Karnataka sample:

$$\begin{aligned} y_{ivt} = \alpha &+ \beta_1 \mathrm{MF}_v \times \mathrm{Post}_t + \gamma_1 \mathrm{MF}_v \times H_{iv} \times \mathrm{Post}_t + \gamma_2 H_{iv} \times \mathrm{Post}_t + \gamma_3 \mathrm{MF}_v \times H_{iv} \\ &+ \delta_1 \mathrm{MF}_v + \delta_2 H_{iv} + \delta_3 \mathrm{Post}_t + \delta' X_{ij,v} + \epsilon_{ivt}, \end{aligned}$$

where again y_{ivt} is the amount borrowed from the stated source (MFI, friends, self-help group, family, moneylenders).

Table 9, Panel A presents the results for the Karnataka data. In column 1, we find that L households do borrow from MFI (i.e. the classification is not perfect); the coefficient is 477 rupees ($p \leq 0.001$). However Hs borrow much more than Ls do (by 1,787 rupees). Columns 2 and 3 find that L households experience a reduction in borrowing from friends and self-help groups (SHGs): they lose Rs. 562 ($p = 0.089$) in loans from friends and Rs. 845 ($p = 0.029$) from SHGs after their village is exposed to microfinance.[41] ,[42]

[41]SHGs are groups of women who get together to lend to each other.

[42]What is striking is that even Ls with no H links lose an enormous amount of borrowing (INR 1294 from friends, INR 1660 from family), and having H friends only mildly affects the total borrowing (and in inconsistent ways across friends and family). See Table G.1 in Appendix G.

Panel B of Table 9 turns to the Hyderabad data. Here we present impacts on borrowing measured in the first post-microfinance survey wave, 15-18 months after microfinance was introduced in treatment neighborhoods.[43]

$$y_{ivt} = \alpha + \beta_1 \mathrm{MF}_v + \gamma_1 \mathrm{MF}_v \times H_{iv} + \delta_2 H_{iv} + \delta' X_{ij,v} + \epsilon_{ivt},$$

Starting with column 1, we find no impact of MF on borrowing for Ls but a large and significantly greater impact for H types (8,776 rupees, $p \leq 0.001$).

We again observe a decline in some types of informal borrowing for L types. Unlike in Karnataka, we detect no change in borrowing from friends. However, we do find large changes in L type borrowing from SHGs (-1,883, $p = 0.021$) and moneylenders (-2,664, $p = 0.071$). Note that H types experience no decline in moneylender borrowing with access to microfinance. In fact they gain Rs. 4,690 ($p = 0.077$) relative to the L types.

While the evidence suggests that markets for informal credit likely function somewhat differently in rural versus urban settings, we do find that exposure to microfinance has an adverse effect on the network borrowing of the Ls in both contexts. This is especially striking because, all else being equal, we would have expected Ls' informal borrowing to go down less than that of the Hs, or even to go up to the extent there is re-lending of microfinance loans (from Hs to Ls).

5.2. **Impact on Risk Sharing.** One important role of village networks is risk sharing, both through gifts and through "soft" loans whose terms are state-contingent (see, e.g., Townsend (1994); Udry (1994)). In light of our finding that access to formal credit results in the loss of network links and a reduction in informal borrowing, it is natural to investigate whether risk sharing is also affected.

We begin by presenting reduced form evidence that the introduction of microfinance worsens risk-sharing for L types and differentially so, relative to H types. The finding on its own is important and striking. That there is a reduction in consumption smoothing among those who are ex-ante unlikely to interact with the treatment means that there are non-obvious externalities that a policymaker or intervening agent must take cognizance of.

Our Hyderabad data contains panel information on both income and consumption[44], which allows us to run standard omnibus tests of risk sharing. Specifically we estimate a version of the standard regression from Townsend (1994), which allows the pass-through of income to consumption to differ by treatment status, differentially for H and L types. The subscript i indexes households, v indexes villages, and t indexes time. Microfinance (i.e., treated) villages are denoted with MF. The term α_i is a household-fixed effect which, as well as controlling

[43]Recall that we only measured networks in the third wave, four years later. However, we collected information about informal borrowing in the earlier waves.

[44]Recall, the Hyderabad data only contain network information collected in 2012, but income and consumption were collected in 2007-8, 2010 and 2012.

for the household's time-invariant Pareto weight in a risk-sharing regression, absorbs the main effect of treatment status and of type (H or L). The term γ_{vt} is a area-time fixed effect which captures the aggregate shock to a particular neighborhood in a given survey wave. We estimate the following "long differences" specification, using the first and third waves of the data.[45]

$$c_{ivt} = \alpha_i + \gamma_{vt} + \beta_1 y_{ivt} + \beta_2 y_{ivt} \times MF_v + \beta_3 y_{ivt} \times H_i + \beta_4 y_{ivt} \times MF_v \times H_i + \delta' X_{ij,v} + \epsilon_{ivt}$$

The coefficient β_1 measures the extent of income pass-though for L households in control areas. The prediction that risk sharing should worsen for L households in treatment, compared to L households in control, is tested by the coefficient β_2. The coefficient β_3 captures the extent of income pass-though for H households in control areas. Finally β_4 measures any differential treatment effect on risk sharing for H households.[46]

Table 10 presents the results. In column 1, we consider per capita non-food consumption, which is typically more discretionary and therefore more responsive to shocks. In column 2, we consider total per capita household expenditure. We find that L households in control areas experience an INR 0.058 drop in non-food consumption for a INR 1 drop in income, a 5.8% pass-through rate ($p = 0.005$). However, for L households in treated areas, the pass-though increases by 0.071 ($p = 0.022$). The estimates are qualitatively similar for total consumption, but slightly less precise ($p = 0.079$ for the hypothesis that pass through is greater in MF areas for L households).

The remaining coefficients in Table 10 consider how these patterns differ for H types. We find suggestive evidence that microfinance causes the pass-through of income into non-food consumption for H types to decrease by 0.065, relative to the effect for L types ($p = 0.153$). For these types, microfinance causes no detectable change to income pass-through ($\beta_2 + \beta_4$, $p = 0.834$).

While inspired by the reduction in links for Ls, we stop short of claiming that this is solely mediated through the reduction in L's links. This is because, of course, there may also be direct effects from microcredit that do not go through the social network. For example, it may be that the microcredit beneficiaries are less willing to contribute to a community fund to help the poor. We have no direct evidence that such a fund exists or that there was any response of

[45]As noted by Hayashi et al. (1996), long differences may perform better than one-period differences if income changes are dominated by transitory measurement error or transitory wage changes, or if income is known one period in advance. Moreover, in our setting, the second wave of data was collected at a point in time when both treatment and control areas had access to microfinance, which may limit our ability to detect treatment effects on risk-sharing. Results using all three waves of data are qualitatively similar but less precise, consistent with the measurement error concern noted by Hayashi et al. (1996).

[46]In Appendix Table L.1, we show that microfinance does not have any impact on income for either L or H types.

this kind, but it cannot be ruled out. In this sense this is part of a more general concern that policy interventions can affect risk sharing relationships.[47]

In summary, these results demonstrate that, while H households' ability to buffer income risk is unaffected by microfinance exposure (or even weakly improved), this is not true for L households. For these households, who are not directly benefiting from the introduction of microcredit, its effects on local networks lead to losses in their ability to smooth risk.

6. Conclusion

By studying the introduction of microcredit in two different settings, we established that not only did the social networks change in response, but those who were least likely to take up microcredit experienced substantial losses in links, even in groups (triangles) where no one was involved in microcredit. This is accompanied by a loss in the ability to borrow from informal networks for those households. The results were first obtained in one setting and then confirmed in a second, independent one. Moreover, in our Hyderabad data, where we can also examine income and consumption fluctuations, we observe a reduction in the ability to smooth risk.

To explain the data, we developed a model in which agents put in effort in order to socialize, whom they meet has an undirected component, and agents engage in mutual consent to build links. Such a model features a global externality, beyond the typical externalities directly embedded in payoffs. When access to microfinance reduces a borrower's desire to maintain and form links with others, even those unlikely to join may reduce their own effort to maintain and build links. This is for two reasons: first, these low types who are linked to potential borrowers consequently have lower returns to such links (decreased payoff externality); second, because of the reduction in overall linking effort, even relationships where the direct payoffs are unaffected by microfinance can be affected. In equilibrium, those who are unlikely to be involved with microcredit may end up with the greatest losses in links.

The fact that our model provides patterns consistent with the data, of course, does not imply that it is the right or only mechanism behind the empirical observations. It will take further research to develop a full understanding of the forces underlying our empirical observations. Nonetheless, the facts – in particular the evidence of negative spillovers on the non-beneficiaries – have wide-ranging and important implications. The previous literature has shown that there may be important benefits from microfinance on participant households beyond the loans themselves, especially in terms of strengthened network connections. But if this comes at a significant cost of weakened connections in the rest of the community, this could worsen the aggregate ability of a community to buffer risk.

[47] Angelucci and De Giorgi (2009), for example, analyze the Progresa cash transfer program in Mexico and find that even non-eligible households consume more. Angelucci et al. (2018) trace these impacts through the kinship network. Albarran and Attanasio (2003) also analyze the interplay between policies and risk sharing, highlighting the potential for risk sharing to go down due to improved outside options.

Regardless of the explanation for the changes, the more general lesson these findings illustrate is that social networks can involve spillovers, externalities, and complex relationships so that changing one part of the network can have quite extensive and unanticipated consequences elsewhere. As a result, interventions into a community can change the social structure and interactions in ways that no one intended, with potentially large costs for some non-participants. Being mindful of these possibilities is important in designing effective policies.

References

Albarran, P. and O. P. Attanasio (2003): "Limited commitment and crowding out of private transfers: evidence from a randomised experiment," *The Economic Journal*, 113, C77–C85.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 113, 151–184.

Ambrus, A., M. Mobius, and A. Szeidl (2014): "Consumption risk-sharing in social networks," *American Economic Review*, 104, 149–82.

Angelucci, M. and G. De Giorgi (2009): "Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption?" *American economic review*, 99, 486–508.

Angelucci, M., G. De Giorgi, and I. Rasul (2018): "Consumption and investment in resource pooling family networks," *The Economic Journal*, 128, 2613–2651.

Angelucci, M., D. Karlan, and J. Zinman (2015): "Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco," *American Economic Journal: Applied Economics*, 7, 151–82.

Arrow, K. J. (2000): "Observations on social capital," *Social capital: A multifaceted perspective (World Bank Publisher)*, 6, 3–5.

Attanasio, O., B. Augsburg, R. De Haas, E. Fitzsimons, and H. Harmgart (2015): "The impacts of microfinance: Evidence from joint-liability lending in Mongolia," *American Economic Journal: Applied Economics*, 7, 90–122.

Augsburg, B., R. De Haas, H. Harmgart, and C. Meghir (2015): "The impacts of microcredit: Evidence from Bosnia and Herzegovina," *American Economic Journal: Applied Economics*, 7, 183–203.

Badev, A. (2013): "Discrete Games in Endogenous Networks: Theory and Policy," *mimeo: University of Pennsylvania*.

Bala, V. and S. Goyal (2000): "A noncooperative model of network formation," *Econometrica*, 68, 1181–1229.

Banerjee, A., E. Breza, E. Duflo, and C. Kinnan (2019a): "Can Microfinance Unlock a Poverty Trap for Some Entrepreneurs?" Tech. rep., National Bureau of Economic Research.

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2019b): "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *The Review of Economic Studies*, 86, 2453–2490.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015a): "The miracle of microfinance? Evidence from a randomized evaluation," *American Economic Journal: Applied Economics*, 7, 22–53.

Banerjee, A., D. Karlan, and J. Zinman (2015b): "Six randomized evaluations of microcredit: Introduction and further steps," *American Economic Journal: Applied Economics*, 7, 1–21.

Beaman, L. and J. Magruder (2012): "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 102, 3574–93.

Belloni, A. and V. Chernozhukov (2009): "Least squares after model selection in high-dimensional sparse models," *MIT Department of Economics Working Paper*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014a): "High-dimensional methods and inference on structural and treatment effects," *The Journal of Economic Perspectives*, 28, 29–50.

——— (2014b): "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 81, 608–650.

Binzel, C., E. Field, and R. Pande (2013): "Does the Arrival of a Formal Financial Institution Alter Informal Sharing Arrangements? Experimental Evidence from Village India," Working paper.

Bloch, F., G. Genicot, and D. Ray (2008): "Informal insurance in social networks," *Journal of Economic Theory*, 143, 36–58.

Blumenstock, J. and X. Tan (2016): "Social Networks and Migration: Theory and Evidence from Rwanda," *Working Paper*.

Blumenstock, J. E., N. Eagle, and M. Fafchamps (2016): "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters," *Journal of Development Economics*, 120, 157–181.

Boucher, V. (2015): "Structural homophily," *International Economic Review*, 56, 235–264.

Breiman, L. and P. Spector (1992): "Submodel selection and evaluation in regression. The X-random case," *International statistical review/revue internationale de Statistique*, 291–319.

Breza, E. (2016): "Field experiments, social networks, and development," *The Oxford handbook of the economics of networks*.

Breza, E. and A. G. Chandrasekhar (2019): "Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment," *Econometrica*, 87, 175–216.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2019): "Consistently estimating graph statistics using Aggregated Relational Data," *arXiv preprint arXiv:1908.09881*.

——— (2020a): "Data and Code for: Using aggregate relational data to feasibly identify network structure without network data," .

——— (2020b): "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 110, 2454–84.

Breza, E. and C. Kinnan (2021): "Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis," *The Quarterly Journal of Economics*, 136, 1447–1497.

Cabrales, A., A. Calvó-Armengol, and Y. Zenou (2011): "Social interactions and spillovers," *Games and Economic Behavior*, 72, 339–360.

Canen, N., M. O. Jackson, and F. Trebbi (2017): "Endogenous Networks and Legislative Activity," *SSRN http://ssrn.com/abstract=2823338*.

Chandrasekhar, A. and M. O. Jackson (2018): "A Network Formation Model Based on Subgraphs," *SSRN Working Paper: http://ssrn.com/abstract=2660381*.

Chandrasekhar, A. and R. Lewis (2014): "Econometrics of sampled networks," MIT working paper.

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010a): "An Empirical Model for Strategic Network Formation," *NBER Working Paper*.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010b): "An empirical model for strategic network formation," Tech. rep., National Bureau of Economic Research.

Coate, S. and M. Ravallion (1993): "Reciprocity without commitment: Characterization and performance of informal insurance arrangements," *Journal of development Economics*, 40, 1–24.

Comola, M. and S. Prina (2017): "Treatment effect accounting for network changes," *Review of Economics and Statistics*, 1–25.

Crépon, B., F. Devoto, E. Duflo, and W. Parienté (2015): "Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco," *American Economic Journal: Applied Economics*, 7, 123–50.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045.

——— (2010): "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences*, 107, 4857–4861.

Currarini, S. and M. Morelli (2000): "Network formation with sequential demands," *Review of Economic Design*, 5, 229 – 250.

Dupas, P., A. Keats, and J. Robinson (2019): "The effect of savings accounts on interpersonal financial relationships: Evidence from a field experiment in rural Kenya," *The*

Economic Journal, 129, 273–310.

Dutta, B. and S. Mutuswami (1997): "Stable networks," *Journal of economic theory*, 76, 322–344.

Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of development Economics*, 71, 261–287.

Feigenberg, B., E. Field, and R. Pande (2013): "The economic returns to social interaction: Experimental evidence from microfinance," *The Review of Economic Studies*, rdt016.

Field, E., R. Pande, J. Papp, and Y. J. Park (2012): "Repayment flexibility can reduce financial stress: a randomized control trial with microfinance clients in India," *PloS one*, 7, e45679.

Hayashi, F., J. Altonji, and L. Kotlikoff (1996): "Risk-Sharing between and within Families," *Econometrica: Journal of the Econometric Society*, 261–294.

Herings, P. J.-J., A. Mauleon, and V. Vannetelbosch (2009): "Farsightedly stable networks," *Games and Economic Behavior*, 67, 526–541.

Hess, S. H., D. Jaimovich, and M. Schündeln (2020): "Development Projects and Economic Networks: Lessons From Rural Gambia," *The Review of Economic Studies*.

Hoff, P., A. Raftery, and M. Handcock (2002): "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97:460, 1090–1098.

Jackson, M. O. (2003): "The stability and efficiency of economic and social networks," *Advances in Economic Design, Heidelberg: Springer-Verlag, edited by Koray, S. and Sertel, M.*

——— (2008): *Social and economic networks*, Princeton: Princeton University Press.

Jackson, M. O., T. R. Rodriguez-Barraquer, and X. Tan (2012): "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *American Economic Review*, 102, 1857–1897.

Jackson, M. O. and A. Van den Nouweland (2005): "Strongly stable networks," *Games and Economic Behavior*, 51, 420–444.

Jackson, M. O. and A. Watts (2002): "The evolution of social and economic networks," *Journal of Economic Theory*, 106(2), 265–295.

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

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009): "Trust and Social Collateral," *The Quarterly Journal of Economics*, 24, 1307–1361.

Kinnan, C. and R. Townsend (2012): "Kinship and financial networks, formal financial access, and risk reduction," *American Economic Review*, 102, 289–93.

König, M., C. Tessone, and Y. Zenou (2014): "Nestedness in networks: A theoretical model and some applications," *Theoretical Economics, forthcoming*.

Kranton, R. E. (1996): "Reciprocal exchange: a self-sustaining system," *The American Economic Review*, 830–851.

Ligon, E., J. P. Thomas, and T. Worrall (2000): "Mutual insurance, individual savings, and limited commitment," *Review of Economic Dynamics*, 3, 216–246.

Macchiavello, R. and A. Morjaria (2021): "Competition and relational contracts in the Rwanda coffee chain," *The Quarterly Journal of Economics*, 136, 1089–1143.

McMillan, J. and C. Woodruff (1999): "Interfirm relationships and informal credit in Vietnam," *The Quarterly Journal of Economics*, 114, 1285–1320.

Meager, R. (2015): "Understanding the impact of microcredit expansions: A bayesian hierarchical analysis of 7 randomised experiments," *arXiv preprint arXiv:1506.06669*.

Mele, A. (2017): "A structural model of Dense Network Formation," *Econometrica*, 85, 825–850.

Munshi, K. and M. Rosenzweig (2016): "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 106, 46–98.

Putnam, R. (2000): *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.

Stone, M. (1974): "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, 36, 111–133.

Tarozzi, A., J. Desai, and K. Johnson (2015): "The impacts of microcredit: Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 7, 54–89.

Townsend, R. M. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539–591.

Udry, C. (1994): "Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria," *The Review of Economic Studies*, 61, 495–526.

Van Zandt, T. and X. Vives (2007): "Monotone equilibria in Bayesian games of strategic complementarities," *Journal of Economic Theory*, 134, 339–360.

Vera-Cossio, D. A. (2019): "Targeting Credit through Community Members," .

Watts, A. (2001): "A dynamic model of network formation," *Games and Economic Behavior*, 34, 331–341.

Xu, Y. and R. Goodacre (2018): "On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," *Journal of analysis and testing*, 2, 249–262.

Tables

Table 1. Sample Statistics

Panel A: Karnataka Wave 1 Households

				Treatment - control		
	Obs	Control Mean	Control SD	Coeff.	5% limit	p-value
Eligible Female	7511	0.943	0.233	0.008	0.015	0.216
No access to latrine	7511	0.748	0.434	-0.038	0.051	0.205
Number of rooms	7511	2.489	1.313	-0.001	0.140	0.973
Thatched roof	7511	0.021	0.145	-0.002	0.014	0.678
Distance to Bangalore	7511	61.114	17.458	-3.823	8.074	0.309
All loans	7218	37861.564	129797.423	1351.740	11597.294	0.819
Network (friends and family) loans	7218	2735.470	25394.731	6.467	1716.401	0.994
SHG loans	7218	2543.994	6944.324	14.783	968.668	0.976
Bank loans	7218	19892.356	106358.225	3563.106	8808.589	0.428
Moneylender loans	7218	3638.339	20456.671	-164.660	1656.949	0.846
Distance to leaders > 1	7511	0.206	0.404	0.042	0.056	0.163
Leader	7511	0.154	0.361	-0.007	0.021	0.555
Share WC leader	7511	0.516	0.294	-0.060	0.078	0.112
No electricity	7511	0.075	0.263	-0.018	0.022	0.134
GMOBC	7511	0.709	0.454	-0.039	0.065	0.221
Num. beds	7511	0.912	1.222	0.022	0.213	0.737
RCC roof	7511	0.147	0.355	-0.031	0.045	0.167
Household size	7511	5.014	2.205	0.297	0.231	0.013
Own rent	7511	0.100	0.300	-0.009	0.044	0.810
Distance to town	7511	5.647	3.595	1.203	1.862	0.190
H (RF)	7511	0.542	0.498	-0.209	0.068	0.000
H (logit)	7511	0.752	0.432	0.005	0.042	0.800

Panel B: Hyderabad

				Treatment - control		
	Obs	Control Mean	Control SD	Coeff.	5% limit	p-value
Total outstanding debt in area, baseline	104	39675.337	47776.778	-6981.245	13853.634	0.326
Area population, baseline	104	264.615	160.467	-3.385	58.378	0.910
Total number of businesses in area, baseline	104	7.288	5.003	-0.346	1.927	0.726
Area mean monthly per-capita exp, baseline	104	1004.974	171.510	42.847	70.733	0.238
Area literacy rate (HH heads), baseline	104	0.625	0.167	0.007	0.056	0.811
Area literacy rate (all), baseline	104	0.687	0.094	0.000	0.032	0.976
Prime-aged (18-45) women in HH, endline 1	6863	1.456	0.820	-0.024	0.056	0.413
Owns land in Hyderabad, endline 1	6863	0.061	0.239	-0.001	0.014	0.897
Owns land in village, endline 1	6863	0.194	0.396	0.006	0.057	0.841
HH had a business pre-intervention	6863	0.308	0.462	0.007	0.042	0.736
HH size (adult equiv), endline 1	6863	4.690	1.784	-0.008	0.130	0.899
Adults (16+) in HH, endline 1	6863	3.887	1.754	-0.018	0.129	0.780
Children (<15) in HH, endline 1	6863	1.738	1.310	-0.014	0.104	0.797
Male head of household, endline 1	6863	0.895	0.307	0.012	0.021	0.266
Age of head of household, endline 1	6863	41.146	10.228	-0.226	0.774	0.566
Head of HH with no education, endline 1	6863	0.312	0.463	0.001	0.044	0.975
Any child 13-18 in HH, endline 1	6863	0.452	0.498	0.016	0.031	0.305
Spouse is literate, endline 1	6863	0.543	0.498	0.003	0.049	0.919
Spouse works for a wage, endline 1	6863	0.234	0.423	-0.020	0.048	0.405
H (RF)	6863	0.241	0.427	-0.016	0.077	0.684
H (logit)	6863	0.236	0.425	0.010	0.050	0.705

Notes: This table presents summary statistics and baseline balance for Karnataka (Panel A) and Hyderabad (Panel B). The "5% limit" column shows how large the difference between treatment and control would have needed to be to be significant at the 5% level. The "p-value" column shows a test of significance on the difference between treatment and control. In Karnataka, there are 75 villages in the sample; 43 received microfinance. In Hyderabad, 104 neighborhoods were subject to randomized assignment of microfinance. GMOBC = A dummy for whether the household is general caste or other backwards caste; the omitted categories are scheduled caste and scheduled tribes. General and OBC are considered upper caste. RCC is Reinforced Cement Concrete. The values labeled "baseline" in Hyderabad are at the neighborhood level, while "endline 1" values are at the household level. $H(RF)$ is the random forest classification into high ($H = 1$) or low ($H = 0$) microfinance propensity. $H(logit)$ is the corresponding logit classification: see section 2.4 for details.

Table 2. Characteristics of H versus L

Panel A: Karnataka - Demographics and Amenities variables

	(1) GMOBC	(2) Latrine	(3) Num. Rooms	(4) Num. Beds	(5) Thatched Roof	(6) RCC Roof
H	-0.221	-0.115	-0.177	-0.213	0.019	-0.053
	(0.021)	(0.018)	(0.042)	(0.038)	(0.005)	(0.008)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Depvar Mean	0.7	0.261	2.36	0.84	0.0235	0.117
Observations	14,904	14,904	14,904	14,904	14,904	14,904

Panel B: Karnataka - Network variables

	(1) Degree	(2) Links to L	(3) Links to H	(4) Eig. Cent.
H	1.947	-0.432	2.355	0.017
	(0.254)	(0.208)	(0.164)	(0.002)
	[0.000]	[0.000]	[0.009]	[0.000]
Depvar Mean	8.97	4.61	3.09	0.0524
Observations	14,904	14,904	14,904	14,904

Panel C: Hyderabad - Demographics and Amenities variables

	(1) GMOBC	(2) Latrine	(3) Num. Rooms	(4) Thatched Roof	(5) RCC Roof
H	0.007	0.041	0.185	0.001	-0.026
	(0.040)	(0.030)	(0.104)	(0.010)	(0.027)
	[0.852]	[0.167]	[0.076]	[0.954]	[0.338]
Observations	4,520	4,483	4,516	4,516	4,508
Depvar Mean	0.429	0.578	2.314	0.025	0.882

Panel D: Hyderabad - Network Variables

	(1) Exp. Degree	(2) Exp. Links to L	(3) Exp. Links to H	(4) Exp. Eig. Cent.
H	0.182	-0.651	0.834	0.009
	(0.149)	(0.194)	(0.215)	(0.005)
	[0.222]	[0.001]	[0.000]	[0.074]
Observations	4,523	4,523	4,523	4,523
Depvar Mean	5.813	4.353	1.463	0.075

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Panels A and B pertains to Karnataka, based on Wave 1 data only. In Panels C and D pertaining to Hyderabad, the estimates reflect H-vs.L differences for the non-microfinance (control group) sample only. GMOBC = A dummy for whether the household consists of general caste or other backwards caste, so the omitted categories are scheduled caste and scheduled tribes. General and OBC are considered upper caste. RCC is Reinforced Cement Concrete.

TABLE 3. Effects of Microfinance on Graph-Level Characteristics

Panel A: Karnataka

	(1) Density	(2) Density	(3) Density	(4) Clustering	(5) Clustering	(6) Clustering	(7) Closeness	(8) Closeness	(9) Closeness
Microfinance × Post	-0.0119	-0.0128	-0.0128	0.00357	0.00968	0.00968	-0.0225	-0.0153	-0.0155
	(0.00678)	(0.00690)	(0.00716)	(0.0146)	(0.0147)	(0.0153)	(0.00970)	(0.0117)	(0.0122)
	[0.0836]	[0.0669]	[0.0769]	[0.807]	[0.513]	[0.528]	[0.0234]	[0.193]	[0.208]
Microfinance	-0.0205	0.00477	0.00204	-0.0408	-0.0179	-0.00638	-0.0129	0.00947	0.00963
	(0.00842)	(0.00555)	(0.00227)	(0.0159)	(0.0148)	(0.00551)	(0.00993)	(0.0101)	(0.0106)
	[0.0175]	[0.393]	[0.373]	[0.0123]	[0.230]	[0.250]	[0.199]	[0.353]	[0.366]
Post	-0.0117	-0.0145	-0.0145	-0.00913	0.00852	0.00852	0.105	-0.0472	-0.00499
	(0.00576)	(0.0107)	(0.0111)	(0.0100)	(0.0249)	(0.0258)	(0.00762)	(0.0522)	(0.0778)
	[0.0454]	[0.182]	[0.198]	[0.366]	[0.733]	[0.742]	[0.000]	[0.369]	[0.949]
Observations	150	150	150	150	150	150	150	150	150
Double-Post LASSO		✓	✓		✓	✓		✓	✓
Village FE			✓			✓			✓
Non MF Mean	0.1135	0.1135	0.1135	0.329	0.329	0.329	0.431	0.431	0.431
Depvar Mean	0.0983	0.0983	0.0983	0.307	0.307	0.307	0.418	0.418	0.418

Panel B: Hyderabad

	(1) Density	(2) Density	(3) Clustering	(4) Clustering	(5) Closeness	(6) Closeness
Microfinance	-0.006	-0.005	-0.006	-0.002	-0.002	-0.001
	(0.003)	(0.003)	(0.007)	(0.007)	(0.002)	(0.002)
	[0.086]	[0.062]	[0.426]	[0.771]	[0.501]	[0.575]
Double-Post LASSO	No	Yes	No	Yes	No	Yes
Depvar Mean	0.0286	0.0286	0.052	0.052	0.00262	0.00262
Non MF Mean	0.0318	0.0318	0.053	0.053	0.00312	0.00312
Observations	4,429	4,429	89	89	89	89

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. For Panel A (Karnataka), controls consist of the share of upper caste households, number of households in the village, share of households in self-help groups, share Hindu, share with a latrine in the house, share that own the household, share that have electricity and share that are leaders. For Panel B (Hyderabad), controls are the same demographic characteristics of households and villages that are used in random forest classification of H vs L. In Panel B, Columns (1) and (2) for density were found through direct elicitation at the node level, whereas columns (3) - (6) for clustering and closeness were estimated at the graph level through the ARD survey.

Table 4. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance	-0.058	-0.059	-0.023	-0.021
	(0.018)	(0.019)	(0.008)	(0.008)
	[0.002]	[0.002]	[0.006]	[0.008]
Microfinance $\times$ LH	0.009	0.001	0.007	0.007
	(0.015)	(0.014)	(0.004)	(0.004)
	[0.573]	[0.935]	[0.120]	[0.109]
Microfinance $\times$ HH	0.039	0.023	0.009	0.012
	(0.022)	(0.022)	(0.007)	(0.007)
	[0.086]	[0.292]	[0.206]	[0.059]
LH	-0.025	-0.005	-0.002	-0.006
	(0.012)	(0.011)	(0.004)	(0.004)
	[0.036]	[0.637]	[0.566]	[0.095]
HH	0.008	0.041	0.021	0.008
	(0.017)	(0.017)	(0.006)	(0.006)
	[0.622]	[0.020]	[0.001]	[0.189]
Observations	57,376	57,376	846,561	846,561
Linked Pre-MF	Yes	Yes	No	No
Controls		✓		✓
Depvar Mean	0.441	0.441	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.0753	0.0753
MF + MF x $LH = 0$ p-val	0.014	0.005	0.015	0.014
MF + MF x $HH = 0$ p-val	0.361	0.088	0.101	0.232
MF + LH x MF = MF + HH x MF p-val	0.137	0.286	0.641	0.245

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

Table 5. Link Evolution, Hyderabad

	(1) Prob. Linked	(2) Prob. Linked
Microfinance	-0.005	-0.007
	(0.002)	(0.002)
	[0.035]	[0.004]
Microfinance x LH	0.002	-0.001
	(0.003)	(0.002)
	[0.577]	[0.764]
Microfinance x HH	-0.011	-0.007
	(0.008)	(0.006)
	[0.203]	[0.281]
LH	0.002	0.003
	(0.003)	(0.002)
	[0.532]	[0.113]
HH	0.018	0.014
	(0.008)	(0.006)
	[0.020]	[0.024]
Observations	141,996	141,996
Controls	No	Yes
Depvar Mean	0.0255	0.0255
LL, Non MF Mean	0.0268	0.0268
MF + MF x $LH = 0$ p-val	0.387	0.019
MF + MF x $HH = 0$ p-val	0.066	0.041
MF + MF x $HH =$ MF + MF x LH p-val	0.038	0.18

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. The controls are selected by double post lasso among all variables that are used for random forest classification.

Table 6. Link Evolution for Information and Financial Links

Panel A: Karnataka

	(1) Financial Linked Post-MF	(2) Financial Linked Post-MF	(3) Info Linked Post-MF	(4) Info Linked Post-MF
Microfinance	−0.052	−0.012	−0.050	−0.015
	(0.021)	(0.005)	(0.018)	(0.005)
	[0.014]	[0.016]	[0.006]	[0.007]
Microfinance × LH	−0.005	0.003	0.002	0.006
	(0.019)	(0.003)	(0.017)	(0.003)
	[0.813]	[0.213]	[0.892]	[0.076]
Microfinance × HH	0.029	0.004	0.040	0.005
	(0.026)	(0.005)	(0.021)	(0.005)
	[0.263]	[0.373]	[0.063]	[0.284]
LH	−0.015	−0.001	−0.014	−0.004
	(0.015)	(0.003)	(0.013)	(0.003)
	[0.331]	[0.623]	[0.288]	[0.183]
HH	−0.00002	0.013	−0.004	0.012
	(0.022)	(0.004)	(0.017)	(0.004)
	[1.000]	[0.003]	[0.813]	[0.005]
Observations	27,072	876,865	37,044	866,893
Linked Pre-MF	Yes	No	Yes	No
Depvar Mean	0.333	0.0341	0.326	0.0377
LL, Non-MF Mean	0.371	0.04	0.361	0.0464
MF + MF x $LH = 0$ p-val	0.005	0.035	0.008	0.025
MF + MF x $HH = 0$ p-val	0.343	0.157	0.615	0.061
MF + LHxMF = MF + HHxMF p-val	0.14	0.828	0.046	0.898

Panel B: Hyderabad

	(1) Number of Financial Links (Degree)	(2) Number of Financial Links (Degree)	(3) Number of Non Financial Links (Degree)	(4) Number of Non Financial Links (Degree)
Microfinance	−0.363	−0.393	−0.195	−0.157
	(0.137)	(0.140)	(0.118)	(0.113)
	[0.010]	[0.006]	[0.101]	[0.169]
Microfinance × H	0.571	0.748	0.651	0.675
	(0.214)	(0.203)	(0.212)	(0.206)
	[0.010]	[0.0004]	[0.003]	[0.002]
H	0.071	−0.195	−0.191	−0.298
	(0.149)	(0.145)	(0.164)	(0.165)
	[0.634]	[0.182]	[0.247]	[0.074]]
Observations	4,429	4,429	4,429	4,429
Double-Post LASSO	No	Yes	No	Yes
Depvar Mean	4.24	4.24	2.87	2.87
MF + MFxH = 0 p-val	0.362	0.090	0.045	0.013

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Panel A: Columns 1-2 restrict to financial links; columns 3-4 restrict to non-financial links. Columns 1 and 3 consider links that existed in Wave 1, while columns 2 and 4 consider pairs of nodes that were not linked in Wave 1.

Panel B: Dependent variables in all columns are node-level self-reported counts of financial vs. non-financial links. All columns include a full set of controls. Centrality controls are a vector of flexible controls (a polynomial) for centrality of both nodes. Household characteristics are caste and wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material. Household predictor variables consist of all variables that are used in the random forest classification. In every case we include interactions of all of these network, demographic, and classification variables with the microfinance indicator.

Table 7. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance	−0.078	−0.070	−0.085	−0.076
	(0.029)	(0.026)	(0.023)	(0.019)
	[0.008]	[0.008]	[0.000]	[0.000]
Microfinance × LLH	0.026	0.015	0.043	0.029
	(0.021)	(0.019)	(0.018)	(0.015)
	[0.228]	[0.437]	[0.015]	[0.050]
Microfinance × LHH	0.054	0.028	0.057	0.031
	(0.030)	(0.025)	(0.025)	(0.018)
	[0.072]	[0.256]	[0.022]	[0.092]
Microfinance × HHH	0.093	0.049	0.087	0.048
	(0.042)	(0.038)	(0.031)	(0.026)
	[0.028]	[0.199]	[0.006]	[0.061]
LLH	−0.024	−0.003	−0.037	−0.019
	(0.018)	(0.017)	(0.014)	(0.013)
	[0.180]	[0.879]	[0.009]	[0.137]
LHH	−0.037	0.009	−0.032	0.002
	(0.025)	(0.023)	(0.017)	(0.014)
	[0.133]	[0.696]	[0.053]	[0.871]
HHH	−0.025	0.042	−0.012	0.034
	(0.033)	(0.029)	(0.022)	(0.020)
	[0.454]	[0.151]	[0.593]	[0.090]
Observations	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes
Controls		✓		✓
Depvar Mean	0.197	0.197	0.808	0.808
LLL, Non-MF Mean	0.252	0.252	0.864	0.864
MF + MF x $HHH = 0$ p-val	0.698	0.549	0.935	0.209
MF + MF x $LLH = 0$ p-val	0.023	0.03	0.022	0.025
MF + MF x $LHH = 0$ p-val	0.262	0.048	0.141	0.018
MF + MF x $HHH =$ MF + MF x LLH p-val	0.076	0.35	0.093	0.459
MF + MF x $HHH =$ MF + MF x LHH p-val	0.212	0.492	0.075	0.307
MF + MF x LLH= MF + MF x LHH p-val	0.122	0.456	0.409	0.934

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

Table 8. Triples Evolution, Hyderabad

All variables x 1000	Full Triangle Linked	Full Triangle Linked
	(1)	(2)
Microfinance	−0.018	−0.034
	(0.010)	(0.020)
	[0.067]	[0.086]
Microfinance × LLH	0.010	−0.012
	(0.011)	(0.013)
	[0.370]	[0.344]
Microfinance × LHH	−0.027	−0.052
	(0.038)	(0.040)
	[0.472]	[0.191]
Microfinance × HHH	−0.177	−0.132
	(0.097)	(0.089)
	[0.067]	[0.139]
LLH	0.000	0.018
	(0.010)	(0.013)
	[0.976]	[0.168]
LHH	0.055	0.072
	(0.036)	(0.042)
	[0.132]	[0.087]
HHH	0.217	0.168
	(0.095)	(0.093)
	[0.023]	[0.071]
Observations	3,341,006	3,341,006
Controls	No	Yes
Depvar Mean	0.0353	0.0353
LLL, Non-MF Mean	0.0359	0.0359
MF + MF × $HHH = 0$ p-val	0.045	0.087
MF + MF × $LLH = 0$ p-val	0.552	0.064
MF + MF × $LHH = 0$ p-val	0.256	0.072
MF + MF × HHH = MF + MF × LLH p-val	0.046	0.144
MF + MF × HHH = MF + MF × LHH p-val	0.041	0.162
MF + MF × LLH = MF + MF × LHH p-val	0.217	0.178

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. The controls are selected by double post lasso among all the variables that are used for its random forest classification, and includes several household and village level characteristics.

Table 9. Borrowing patterns

Panel A: Borrowing Patterns, Karnataka

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance × Post	476.572	−562.308	−844.524	704.391	677.970
	(148.808)	(330.341)	(384.839)	(800.168)	(659.590)
	[0.002]	[0.089]	[0.029]	[0.379]	[0.305]
Microfinance × Post × H	1,795.233	203.926	48.466	−2,210.964	−1,608.814
	(245.414)	(242.383)	(346.884)	(943.562)	(1,185.489)
	[0.000]	[0.401]	[0.889]	[0.020]	[0.175]
Microfinance × H	−0.542	−65.457	232.443	206.495	1,088.834
	(58.782)	(63.966)	(254.356)	(497.672)	(885.048)
	[0.993]	[0.307]	[0.361]	[0.679]	[0.219]
Post × H	189.508	−410.031	91.263	1,828.811	400.044
	(108.311)	(199.678)	(279.643)	(734.643)	(522.679)
	[0.081]	[0.041]	[0.745]	[0.013]	[0.445]
Observations	28,062	27,194	28,062	28,062	28,062
Depvar Mean	596.976	860.228	1863.324	2667.56	1656.881
L, Non-MF Mean	189.671	1148.705	1920.918	2344.905	1711.001
MF x Post x H + MF x Post =0 p-val	0.000	0.255	0.119	0.084	0.325

Panel B: Borrowing Patterns, Hyderabad

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance	−209.748	86.742	−1,882.840	−2,664.192	−256.318
	(235.127)	(894.331)	(801.110)	(1,455.603)	(656.431)
	[0.375]	[0.923]	[0.021]	[0.071]	[0.697]
Microfinance × H	8,312.670	−637.232	−1,577.128	4,689.554	1,796.860
	(448.982)	(1,491.449)	(1,369.064)	(2,622.331)	(1,366.622)
	[0.000]	[0.671]	[0.252]	[0.077]	[0.192]
H	−108.232	−1,792.590	1,251.211	−198.590	−507.290
	(296.017)	(1,293.944)	(1,163.829)	(1,899.306)	(985.862)
	[0.716]	[0.169]	[0.285]	[0.917]	[0.608]
Observations	6,811	6,863	6,863	6,863	6,863
Depvar Mean	3107.86	7895.05	6935.66	18805.06	2620.97
L, Non MF Mean	2091.75	8110.94	7064.44	19601.47	2704.03
MF + MF × H = 0 p-val	0.000	0.664	0.012	0.426	0.245

Notes: Table presents the effect of microfinance access on the loan amounts borrowed from various sources; outcomes are winsorized at the 2.5% level. Panel A pertains to Karnataka and tracks loan amounts from microfinance institutions, friends, family, banks and moneylenders. All of its columns control for surveyed in wave 1 fixed effects. The average rate of inflation over the period between waves in Karnataka was 8% meaning a total of 65%. Panel B pertains to Hyderabad. Outcomes are measured in the first survey wave (2007-08). Here all specifications include demographic household and village controls (those used in random forest classification of H vs L) subject to double-post LASSO. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. MFI: Microfinance Institution; SHG: Self-Help Group

TABLE 10. Risk sharing, Hyderabad

	(1) Expenditures: Non-Food	(2) Expenditures: Total
Microfinance×Income	0.071	0.066
	(0.030)	(0.037)
	[0.022]	[0.079]
Microfinance×Income×H	−0.065	−0.112
	(0.044)	(0.058)
	[0.153]	[0.070]
Household Income per capita	0.058	0.109
	(0.019)	(0.024)
	[0.004]	[0.000]
Household Income per capita×H	0.020	0.076
	(0.025)	(0.043)
	[0.438]	[0.082]
Observations	10,502	10,590
Depvar Mean	1193	2040
L, Non-MF Depvar Mean	1187	2049
Income Mean	1440	1437
L, Non-MF Income Mean	1437	1435
Test: MF x Income + MF x Income x H = 0	0.834	0.407

Notes: Income is total household, monthly per capita earnings from employment or business activities, excluding private and government transfers. Dependent variable is monthly per capita household expenditure. In col. 1, expenditure excludes food and in col. 2, we present non-food expenditure. Data is from the first (2007-08) and third (2012) waves of the Hyderabad survey. Regression includes controls for household fixed effects and wave-by-neighborhood-by-type fixed effects. Additional controls are selected by double post lasso from the set of variables used in the prediction exercise, interacted with type. Standard errors (clustered at the neighborhood level) are reported in parentheses. p-values are reported in brackets.

Appendix A. Proof of Proposition 1

We show there is a unique equilibrium and characterize it, here letting each agent's utility be fully dependent upon their label i.

From our discussion above, it follows directly that a best response must satisfy[48]

$$e_i = \min\left\{1,\ \frac{1}{c_i}\left(u_i + \sum_{j\neq i} \mathrm{E}^+[v_{ij}]\,(1-F(-v_{ij}))\,(1-F(-v_{ij}))\,e_j\right)\right\}.$$

Given the bound that $e_j \leq 1$, and the fact that $u_i > 0$, it follows that for sufficiently large c_i,

$$e_i = \frac{1}{c_i}\left(u_i + \sum_{j\neq i} \mathrm{E}^+[v_{ij}]\,(1-F(-v_{ij}))\,(1-F(-v_{ij}))\,e_j\right),$$

and is strictly between 0 and 1.

Thus, taking c_i to be sufficiently large for each i, we let u be the n-dimensional vector with entries $\frac{1}{c_i}u_i$ and E be the $n \times n$ matrix with ij entries

$$\frac{1}{c_i}\mathrm{E}^+[v_{ij}]\,(1-F(-v_{ij}))\,(1-F(-v_{ij}))$$

Then the characterization of equilibria can be written as

$$e = u + Ee,$$

which has a (unique) solution of $e = (I-E)^{-1}u$, given that E has non-negative values that are less than 1 and so $(I-E)$ is invertible.

Note that two agents of the same type take the same effort by the symmetry of the expected utility in type and uniqueness of equilibrium overall.

Rewriting u to be the $|\Theta|$-dimensional vector with entries $\frac{1}{c_\theta}u_\theta$ and E to be the $|\Theta| \times |\Theta|$ matrix with θ, θ' entries

$$\frac{1}{c_\theta}\mathrm{E}^+[v_{\theta\theta'}]n_{\theta\theta'}\,(1-F(-v_{\theta\theta'}))\,(1-F(-v_{\theta'\theta}))$$

the unique equilibrium is given by

$$e = (I-E)^{-1}u.$$

The result on the comparative statics follows from Proposition 16 in Van Zandt and Vives (2007), noting the strict monotonicity of the best responses in the payoffs and actions of others and the interiority of the equilibrium.

[48]This drops the $n_{\theta\theta'}$ terms, but one can include an indicator n_{ij} and nothing in the argument below changes.

Appendix B. Random Forest Model Description

We use a random forest algorithm (implemented in R) to classify our respondents into two types: those that have a high probability of taking up microfinance loans (H) and those that have a low probability (L), when offered.

B.1. **Algorithm Inputs.**

Input Data:

- N = Set of respondents from all villages,
- N_{mf} = Set of respondents from microfinance villages,
- Y_i = Loan take-up binary outcome for each $i \in N_{mf}$,
- X_i = Set of predictor variables for each $i \in N_{mf}$.

Algorithm Parameters:

- T = Set of trees to grow,
- p = Total number of predictors,
- m = Number of predictors selected at each split,
- c = Cut-off: a vector of length 2 (the winning class for an observation is the one with the maximum ratio of proportion of votes to cut-off),
- t = Fraction of sample to be used as training dataset.

B.2. **Basic Algorithm.**

Step 1: Randomly select (with replacement) training data S and testing data S' from N_{mf}. The size of S will be $t \cdot n(N_{mf})$ and the size of S' will be $(1-t) \cdot n(N_{mf})$.

Step 2: For each tree $t \in T$,
- Randomly select (without replacement) a sample of size $n(S)$ from S.
- At each node n of the tree t, randomly select (with replacement) a set of predictors of size m from p.
- At each node, construct a split based on a rule which uses Gini's Diversity Index (gdi) to determine the split.
- For every tree t, each $i \in N_{mf}$ will be assigned a classification $\hat{Y}_{it} \in \{0, 1\}$.

Step 3: After classifying each $i \in N_{mf}$, for each tree t, the final classification can be computed as follows,

$$\hat{Y}_i = 1\left\{\frac{1}{n(S)}\sum_{t=1}^{n(T)} \hat{Y}_{it} > c[2]\right\}$$

and therefore $\theta_i = \hat{Y}_i \cdot H + (1 - \hat{Y}_i) \cdot L$.

B.3. **Our Parameter Choices.**

- T: We use 1500 trees.
- p: We use 13 predictors for Karnataka and 19 predictors for Hyderabad. The choice of predictors is explained in subsection B.4.

- m: We use the basic R `randomForest` parameter which is equal to $\sqrt{p}$ for classification.
- c: We use the vector $(0.85, 0.15)$ for Karnataka panel and $(0.73, 0.27)$ for Hyderabad panel, chosen by cross-validation.
- t: We use 0.7 of our sample to train the data.

B.4. **Selection of predictors.** Where possible, in both settings, we select predictor variables that are likely correlated with microfinance eligibility, awareness and take-up.

B.4.1. *Karnataka predictors.* In the Karnataka sample, we have detailed information on how the MFI marketed its product to potential borrowers, along with the eligibility rules. The first five predictors capture key components of eligibility and awareness. The network measures are included to pick up the likelihood of households hearing about the product. We also include an additional set of household characteristics associated with wealth.

- dummy for being a BSS leader, who are the people that the MFI would approach when entering a village (the BSS definition of leader was defined by occupation, e.g., teachers; self-help group leaders; shopkeepers, so we can identify them similarly in MF and non-MF villages),
- dummy for whether the household has a female of eligible age (18-57) for a microfinance loan, which is a requirement for the household to be able to participate,
- the average closeness (mean of inverse of network distance) to leaders, which is relevant because those who are closer to leaders are more likely to hear of microfinance (Banerjee et al., 2013),
- the average closeness (mean of inverse distance) to same-caste leaders, because interactions within-caste are more likely and therefore should influence the likelihood of being informed,
- the share of same-caste leaders in the village, as above.
- GMOBC= a dummy for whether the household consists of general caste or other backward caste, so the omitted categories are scheduled caste and scheduled tribes (SCST); general and OBC are considered upper caste,
- household size,
- number of rooms,
- number of beds,
- dummy for access to electricity,
- dummy for access to latrine,
- dummy for RCC roof (considered a superior type of roof),
- dummy for thatched roof.

Table 2 shows a comparison of the H respondents and the L respondents along key dimensions. We see that H households are much more likely to be SCST, have smaller houses in terms of room count, are much less likely to have a latrine in the household, and are much less likely to have an RCC roof, all of which suggests that they tend to be poorer. Finally, we see that H households and L households have comparable degree (H types have 1.94 more friends on a base of 8.97), but the composition exhibits considerable homophily: H types have a lower number of links to L types and a higher number of links to H types. But H households are more eigenvector central in the network.

B.4.2. *Hyderabad predictors.* In Hyderabad, we have less precise information about the lending strategy followed by the local MFI. To be eligible, households needed to have a prime-aged woman, a key variable that we include as a predictor. We also include several neighborhood and household level variables that may proxy for credit supply, credit demand, demographic composition, and proxies for wealth.

- Total outstanding debt in area (baseline)
- Area population (baseline)
- Total number of businesses in area (baseline)
- Area mean monthly per-capita expenditure (baseline)
- Area literacy rate (HH heads only, baseline)
- Area literacy rate (all adults, baseline)
- Dummy for household operating any business(es) prior to 2006, when Spandana opened branches in treatment areas (endline 1)
- Adult equivalent household size (endline 1)
- Adults (16 and older) in household (endline 1)
- Children (15 and younger) in household (endline 1)
- Dummy for male household head (endline 1)
- Age of head of household (endline 1)
- Head of household with no education (endline 1)
- Prime-aged (18-45) women in household (endline 1)
- Any child 13-18 in household (endline 1)
- Dummy for literate spouse of household head (endline 1)
- Dummy for spouse who works for a wage (endline 1)
- Dummy for household owns land in Hyderabad (endline 1)
- Dummy for household owns land in native village (endline 1)

Note: All baseline variables were collected in 2005 and are area-level averages (Baseline households were not systematically resurveyed at endline). Endline 1 variables were collected in 2007-08 and are at the household level. The endline variables were selected to only include pre-determined household characteristics. For more information on the baseline and endline surveys, see Banerjee et al. (2015a).

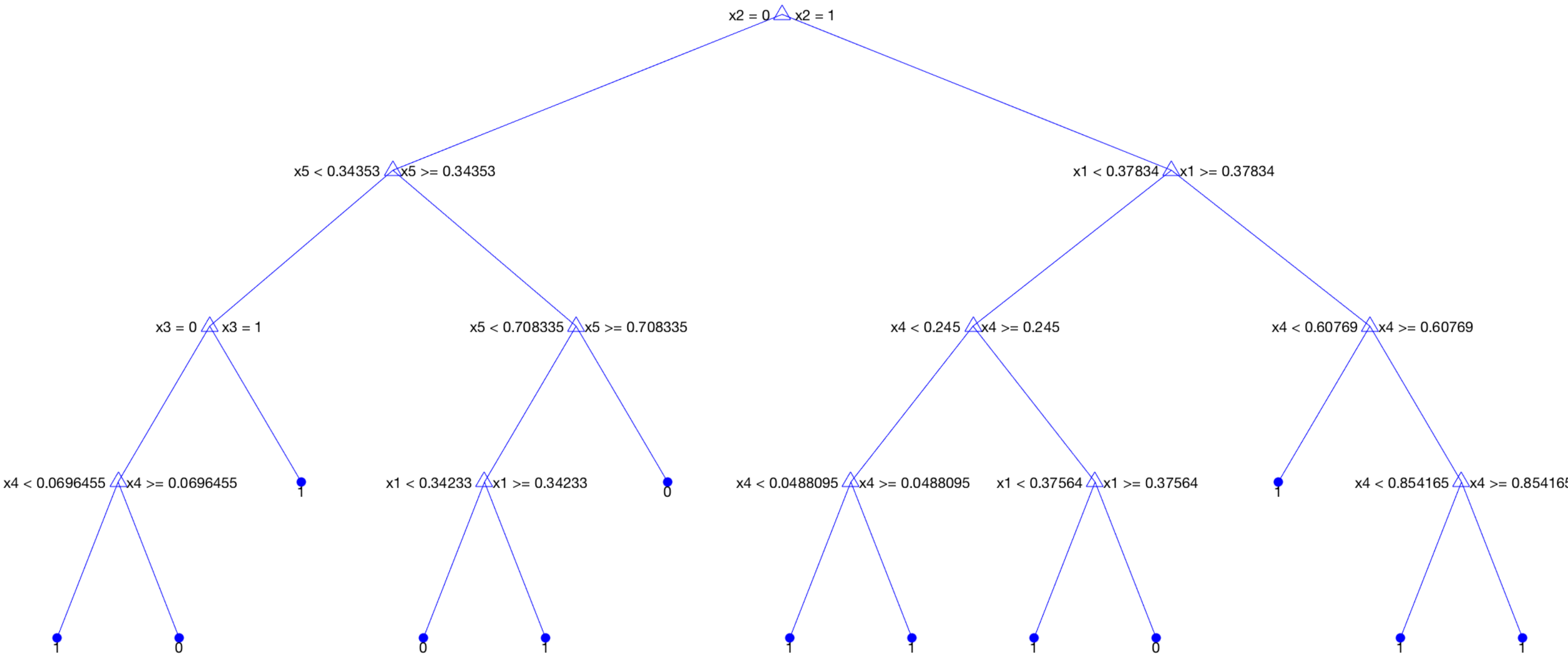


Figure B.1. This presents an example of a decision tree. For the sake of simplicity, we limit the maximum number of splits to 12. The actual procedure has a considerably more complex tree. Here x_1 is the average closeness to leaders, x_2 is whether the household is eligible by having a female of eligible age, x_3 is whether the household is a leader, x_4 is the share of same-caste leaders in the village, and x_5 is the closeness to same-caste leaders.

Table B.1. Confusion Matrices for H and L classification, Karnataka

		Predicted		
		L	H	Total
Observed	L	1469	898	2367
	H	204	308	512
	Total	1673	1206	$N = 2879$

Notes: This table presents the confusion matrix for the validation sample for Karnataka. The following metrics on this confusion matrix capture classification quality: DOR = 2.47, F1 = 0.359, MCC = 0.172.

Table B.2. Confusion Matrices for H and L classification, Hyderabad

		Predicted		
		L	H	Total
Observed	L	661	174	835
	H	129	105	234
	Total	790	279	$N = 1069$

Notes: This table presents the confusion matrix for the validation sample for Hyderabad. The following metrics on this confusion matrix capture classification quality: DOR = 3.09, F1 = 0.409, MCC = 0.226.

B.5. **Random Forest Classifier quality metrics and comparison with Logistic Classifier.** Here we compare the performance of the random forest and logistic classifiers. Appendix Tables B.1 and B.2 present the confusion matrices for the random forest classifiers in Karnataka and Hyderabad, respectively. Appendix Tables J.6 and J.12 present the confusion matrices for the logistic classifiers in both samples.

The confusion matrices present the fractions of true positives, true negatives, false positives and false negatives. Ideally, the true positives and negatives would be 100% each and the rate of false positives and negatives would be 0%. In each of the table notes, we also present several commonly-used diagnostic measurements for assessing the quality of classification. These include Matthews correlation coefficient (the preferred diagnostic measure), the F1 score (TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative), and the the diagnostic odds ratio:

- Matthews correlation coefficient: $MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
- F1 score: $F1 = 2\frac{PPV.TPR}{PPV+TPR}$ where $PPV = \frac{TP}{TP+FP}$ and $TPR = \frac{TP}{TP+FN}$
- Diagnostic odds ratio: $DOR = \frac{TP.TN}{FP.FN}$

Both random forest and logistic classifiers have a cut-off parameter that was chosen by 3-fold cross-validation to maximize the Matthews correlation coefficient. We next compare each classification metric across random forest and logistic. Overall, random forest outperforms logit. The difference is particularly stark in the Hyderabad sample.

B.5.1. *Karnataka Sample.* We compare the quality of the classification for Karnataka:

(1) Confusion Matrix
  (a) True negative rate (TNR):
    - Random forest: 62%
    - Logit: 37%
  (b) True positive rate (TPR):
    - Random forest: 60%
    - Logit: 84%
  (c) Positive predicted value (PPV, probability of true positive vs all positive):
    - Random forest: 26%
    - Logit: 22%
  (d) Negative predicted value (NPV, probability of true negative vs all negative):
    - Random forest: 88%
    - Logit: 91%
(2) Metrics for quality of classification are:
  (a) Matthews correlation coefficient (from [-1,1]): the preferred diagnostic measure
    - Random forest: 0.172
    - Logit: 0.164

(b) F1 score (from $[0,1]$)
  - Random forest: 0.359
  - Logit: 0.351

(c) Diagnostic odds ratio (positive odds ratio / negative odds ratio, from $[0,\infty)$):
  - Random forest: 2.47
  - Logit: 2.94

B.5.2. *Hyderabad Sample.* We compare the quality of the classification for Hyderabad:

(1) Confusion Matrix
  (a) True negative rate:
    - Random forest: 79%
    - Logit: 78%
  (b) True positive rate:
    - Random forest: 45%
    - Logit: 35%
  (c) Positive predicted value (probability of true positive vs all positive):
    - Random forest: 38%
    - Logit: 31%
  (d) Negative predicted value (probability of true negative vs all negative):
    - Random forest: 84%
    - Logit: 81%

(2) Metrics for quality of classification are:
  (a) Matthews correlation coefficient (from [-1,1]): the preferred diagnostic measure
    - Random forest: 0.226
    - Logit: 0.120
  (b) F1 score (from $[0,1]$)
    - Random forest: 0.409
    - Logit: 0.325
  (c) Diagnostic odds ratio (positive odds ratio / negative odds ratio, from $[0,\infty)$):
    - Random forest: 3.09
    - Logit: 1.87

Appendix C. Balance

Table C.1. Covariate balance

Karnataka Wave 1 Villages

							Number of Households Adjusted		
				Treatment - control			Treatment - control		
	Obs	Control Mean	Control SD	Coeff.	p-value	5% limit	Coeff.	p-value	5% limit
Number of Households	75	165.812	48.945	57.397	0.000	24.354			
Average Density	75	0.119	0.042	-0.020	0.011	0.015	0.005	0.467	0.013
Average Clustering	75	0.334	0.074	-0.041	0.009	0.030	-0.018	0.284	0.032
Average Closeness	75	0.379	0.046	-0.013	0.183	0.019	0.009	0.320	0.019
Harm mean distance to leaders	7511	0.474	0.079	-0.021	0.049	0.020	0.007	0.481	0.020
WC harm mean dist	7511	0.495	0.186	-0.020	0.127	0.028	-0.004	0.761	0.028

Notes: This table reports Treatment-Control balance on additional variables used as predictors to classify households as H (high MF propensity) vs. L (low MF propensity). 5% limit shows the size of the difference between treatment and control that would be powered to detect at the 5% level, holding the standard error fixed. Unit of observation: household (except for area variables). p-values of differences reflect standard errors clustered at the area level.

Table C.2. Endline network summary statistics

	Non-Microfinance villages
Panel A: Karnataka Wave 2 Data	
Average Degree (Mean)	17.46
Average Degree (Std. Dev.)	4.34
Average Clustering (Mean)	0.32
Average Clustering (Std. Dev.)	0.06
Average Closeness (Mean)	0.48
Average Closeness (Std. Dev.)	0.05
Number of Households (Mean)	175.84
Number of Households (Std. Dev.)	53.49
Panel B: Hyderabad Data	
Average Degree (Mean)	5.949
Average Degree (Std. Dev.)	0.833
Average Clustering (Mean)	0.064
Average Clustering (Std. Dev.)	0.034
Average Closeness (Mean)	0.004
Average Closeness (Std. Dev.)	0.013
Number of Households (Mean)	200.738
Number of Households (Std. Dev.)	101.377

Appendix D. Alternative Models: Existing Models in the Literature

In this section we describe several alternative models, emphasizing why they, in their basic forms, cannot generate the patterns in our data. We also describe an extension of our model that includes direct payoff externalities across links. While richer environments with components from these alternative models may be able to rationalize our empirical findings, we believe our model offers a new, parsimonious, empirically plausible, and valuable theoretical perspective by considering *global* externalities in network formation. In particular, our model places a central focus on the fact that a change, such as the introduction of a new formal financial product, can have effects even for those who do not adopt it and who may be arbitrarily distant from those who do adopt, by affecting the equilibrium incentives to engage in link formation.

We study four specific alternatives and work through each model using the setup of Section 4.6. What follows, of course, is not an exhaustive list, but is representative of the types of models that would be natural candidates for this application.

The first two involve exogenous random matching and mutual consent. These are analogous to the type of models studied by Watts (2001); Jackson and Watts (2002); Christakis, Fowler, Imbens, and Kalyanaraman (2010b); Mele (2017), albeit presented in a simplified manner for clarity of argument.

First, Section D.1, presents the case when links are historically given but may break as a result of a shock, such as the introduction of microfinance. New links are however slow to form, and, in the short run, the dominant effect of shock is that links break (in the longer run new links presumably form). This is as in Jackson et al. (2012). The second model takes on the opposite extreme case where links get renewed every period from scratch. So in section D.2, we imagine an exogenous set of unlinked individuals who form new links, with random matching opportunities and mutual consent for link formation.

The third model, presented in Section D.3, returns to the case where links are easy to break but slow to form, but focuses on triads rather than pairs. This introduces the idea of support–that the presence of one link may help sustain other links involving some of the same set of people (Jackson et al., 2012).

Despite their very different perspectives, these three models all point to similar conclusions: that the number of HL links should go down in microfinance villages, while the number of LL links should stay the same or, if it does decline, should decline less than mixed link types. Further LLL triads should decline less than LLH or LHH.

The fourth model, presented in Section D.4, returns to the setup where networks essentially re-form every period, but now introduces "directed search". With directed search, agents are free to choose which other types of agents they want to link with. In such a model, we find that while HL links should decline in microfinance villages, LL links should go up. This fits with the main strand of the network formation literature (e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami (1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den

Nouweland (2005); Bloch, Genicot, and Ray (2008); Herings, Mauleon, and Vannetelbosch (2009); Jackson, Rodriguez-Barraquer, and Tan (2012); Boucher (2015)).

Our conclusion is that these four approaches, based on either exogenous or directed search, with mutual consent and perhaps support requirements for forming links, cannot, at least in their basic versions, generate patterns consistent with the data.

D.1. **The impact on pre-existing links.** The first model takes the view that villagers are in a pre-existing network, and while links are easy to break, forming new links can be very slow and is thus not on the same time-scale. We start from a setting where we take these network connections as given before the arrival of microcredit. Where microcredit arrives, people have the choice of continuing or breaking off those relationships, and breaking is unilateral (consistent with mutual consent models). In control villages we assume that nothing changes.

Let us write that the payoff to node i of type θ_i of being linked to j of type θ_j is given by

$$\alpha_{\theta_i}\beta_{\theta_j}r + \beta_{\theta_i}\alpha_{\theta_j}b - \epsilon_{ij},$$

where G is the CDF of ϵ, a mean-zero random variable, so as before the expected value is

$$v_{\theta\theta'} = \alpha_\theta\beta_{\theta'}r + \beta_\theta\alpha_{\theta'}b.$$

Recall that α_θ is the probability of having money to lend and β_θ is the probability of needing to borrow. So we may imagine that, due to microfinance entry, α_H declines with high frequency repayments (or may increase in the world of relending, which appears to be empirically less common) and β_H increases where the microcredit loans allow the borrower to overcome a non-convexity (or may decline if microcredit loans are substitutes for informal loans). The parameters for Ls, who do not borrow from microfinance, remain the same. As a consequence, we imagine that v_{HH} ought to decline and v_{LL} ought to not change. Whether v_{LH} and v_{HL} is a more delicate matter and in general ambiguous.

With this in mind, what is the effect on the number of relationships of each type: HH, LH, and LL? Clearly the number of HH relations goes down and the number of LL relationships should be unchanged. The number of HL relationships however depends on both the willingness of the H to partner with an L, which has gone down and the willingness of an L to partner with an H, which might have gone up. The number of LH pairs in MF villages is given by

$$G(v_{HL} + \Delta v_{HL}) \cdot G(v_{LH} + \Delta v_{LH})$$

compared to

$$G(v_{HL}) \cdot G(v_{LH})$$

in non-MF villages. For relatively small changes in the value of the relationships the difference in the number of HL pairs can be written as

$$G'(v_{HL})\Delta v_{HL} + G'(v_{LH})\Delta v_{LH} = G'(v_{HL})[\Delta v_{HL} + \Delta v_{LH}]$$

$$= (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r + b) < 0.$$

The last inequality follows from the fact that if relending is small relative to the change in appetite for borrowing (as is the case in the empirical literature), then $\Delta_{HH} < 0$, which is the same condition as above.

Therefore the number of HL relations must also fall. Only the number LL relationships do not go down when MF arrives.

Claim **1.** *Starting with a given set of links, the introduction of microfinance should*

(1) *reduce HH links,*
(2) *reduce LH links,*
(3) *leave LL links unchanged,*
(4) *and the total number of links should decline and be less than in non-microfinance villages.*

D.2. **Introducing link formation.** We now turn to a model at the other extreme: there is no persistence in links whatsoever, so we can consider the formation of new links from an unmatched population.

As before the pairs are formed if both parties want the link, which happens with probability $G(v_{\theta\widetilde{\theta}}) \cdot G(v_{\widetilde{\theta}\theta})$ for a $\theta\widetilde{\theta}$ link. From above, the fraction of new HH and LH links should go down in microfinance villages but that of new LL links should remain the same.

Claim **2.** *If new links are formed by randomly matching, the introduction of microfinance should*

(1) *reduce new HH links,*
(2) *reduce new LH links,*
(3) *leave new LL links unchanged,*
(4) *and the total number of new links should be less than in non-microfinance villages.*

D.3. **A model with supported links.** Our third model again takes the perspective that links are easy to break but slow to form, but in this case we focus on the value of a link being supported in the sense of Jackson et al. (2012). Jackson, Rodriguez-Barraquer, and Tan (2012) introduce the notion of support, which correlates the presence of links based on incentives to exchange favors (including lending to each other).The idea is that two households in isolation may not have enough bilateral interaction to be able to sustain cooperation with each other, but if they both also have relationships with some other households in common, then the relationships can all "support" each other: if someone fails to cooperate with one of their friends then beyond losing that relationship, they also lose relationships with all the other friends that they had in common with the friend with whom they did not cooperate. Fear of losing all of those relationships if they misbehave provides added incentives to maintain cooperation.[49] This leads relationships to be correlated: forming them in supported combinations provides stronger incentives, and then both their presence and disappearance ends up being correlated.

This model builds a natural connection between what happens to the Hs (who are directly affected by microcredit) and what happens to Ls. An LL link can break because it is no longer supported by an H. However, for reasons that will become clear, it cannot explain the patterns we observe in the data.

D.3.1. *Payoffs.* We start with a set of HH, LH, and LL links. However some of these links also support each other in the sense that some are part of HHH, LHH, LLH, or LLL triangles. We assume that no one has more than two links to keep the problem manageable. We assume that the payoff to i from the links between i (a type θ) and j (a type $\widetilde{\theta}$) that is supported by k (a type θ') is given by

$$W_{ijk}(\theta, \widetilde{\theta}|\theta') = v_{\theta\widetilde{\theta}} + \max\{\varepsilon_{ij}, \varepsilon_{ik}\}$$

where $v_{\theta\widetilde{\theta}}$ is defined as in Section 4 and ε_{ij} and ε_{ik} are drawn, as before, i.i.d. from a distribution G.

This formulation makes sense in a world where there is no crowd-out in borrowing or lending – when an agent is in the borrowing state he gets twice the benefit b if he can borrow from two sources and when he is in the lending state he gets twice the benefit r if he can lend to two people.

When the relation is not supported, i.e., there is either just one pair or there is a potential triad but not all 3 pairs are connected, the payoff from it is, as before

$$W_{ij}(\theta, \widetilde{\theta}|\emptyset) = v_{\theta\widetilde{\theta}} + \varepsilon_{ij}$$

where the ε_{ij} is drawn, as before, i.i.d. from a distribution G.

D.3.2. *Analysis of the model.* The decision to be made is simple: whether to stay linked. However starting from a trilateral relationship, there are potentially multiple equilibria: i might leave because she expects k to leave and vice versa. To reduce the number of cases, assume that the equilibrium selection rule is always to choose the triad equilibrium if it existed in the pre-period and is still an equilibrium. In other words, each participant of triad only checks whether they want to stay in the relationship if the other two members of the triad were to stay. If the triad is no longer an equilibrium, then each pair in the erstwhile triad independently decides whether or not to stay together as a pair (and clearly at least one will not), and the equilibrium is unique.

[49]This setup also nests risk sharing, which can be seen as another form of favor exchange in which i gives a transfer to j if j is hit by a negative shock. The third friend, k, can be valued for two (non mutually exclusive) reasons: because k will punish i if i reneges on her expected transfer to j and/or because k can make a transfer to i which can then be shared in turn with j.

Clearly some of the Hs who are in a triad and have access to microfinance will want to break at least one link since both v_{HH} and v_{HL} decline. Once this is taken as given, the value of each remaining relationship goes down at least weakly, and in some fraction of cases those relationships will also break up because they were sustained by the higher ε associated with the triad. The only triads that will be unaffected are the LLL triads. All other types of triads will break up more in MF villages than in non-MF villages. It is also easy to see that LHH triads are more likely to break up than LLH triads with microfinance, simply because the LH links are the vulnerable points.

This model can explain why lots of pre-existing LL links break up in MF villages. The argument would be that most of these links were part of a triad with an H and that the H has less incentive to continue in the triad. It does however suggest that fewer LL links should break up than LH links, since under this theory LL links only break up because an LH link that sustained that LL relationship broke up.

Claim **3.** *In the model with supported links, when microfinance is introduced,*

(1) *LLs decline but LHs should decline by more,*
(2) *LHHs are more likely to decline than LLHs, which are more likely to decline than LLLs.*

D.3.3. *Simulation.* To make this transparent, we present a simulation exercise. We look at networks of size $n = 300$. We set the payoff parameters $r = 0.1$, $b = 1$, and $\alpha_H = \alpha_L = \beta_H = \beta_L = 1/3$. We set $\alpha'_H = 1.45\alpha_H$ and vary the needing to borrow probability under microfinance, $\beta'_H \in \{0.25, 0.3, \ldots, 0.65\}$, for the simulations. Under these parameters we have v_{HH}, v_{HL}, v_{LH}, and v_{LL} satisfying the assumptions maintained throughout this paper, described in Section 4. We let $G(\varepsilon) = \mathcal{N}(0, 1/100)$ and let half the population be H and the other half be L.

We repeat 100 simulations of the following procedure. We seed the graph by connecting collections of mutually exclusive sets of three nodes at random. We then draw ε_{ij} and compute an equilibrium network under no-MF payoffs and an equilibrium network under MF payoffs, holding fixed the seed and the shocks as above. Specifically, any triangle that exists initially and for which it is still an equilibrium under the shocks and payoff parameters to maintain are maintained. If not, then constituent links are checked. A resulting equilibrium graph holding fixed seeds and shocks can be computed for each simulation drawn under both non-MF and MF payoffs.

Figure D.1 presents the results. We plot the change in the number of links (and the change in the number of triangles) comparing MF networks to non-MF networks. We see that MF networks uniformly lead to a decline in every link and triangle type. Furthermore, the gap between the models declines the closer β'_H is to β_H. Nonetheless, what is striking is that LL links drop much less than its counterparts HH and LH, as do LLL triangles compared to HHH, LLH and LHH.

D.3.4. *Summary so far.* The models discussed so far, with or without the idea of support, all point to the same conclusion: the number of LH and HH links should go down faster in MF villages than the number of LL links. Moreover LLL triads should be least affected. This is inconsistent with our empirical findings, suggesting that, at least in their basic versions, these models cannot be the (sole) data generating process in our settings.

There is however one additional factor that the previous models ignore. As Feigenberg et al. (2013) show, empirically, microfinance itself may promote connections between group members, who will, by definition, tend to be Hs. This could lead to offsetting effects on HH links and HHH triads, making net predictions ambiguous. We next consider a model of directed search which accommodates this possibility.

D.4. **A model of directed search.** Let us take the set up of the model where networks essentially re-form every period, but now introduce directed search. Instead of matching randomly, we now assume that each agent can select the population within which they will match. Once they observe who they are matched to, which happens randomly within the group, they get to decide whether they will actually form a link. Link formation is unilateral. There are three possible populations: HH (i.e., just Hs), LL (i.e., just Ls), and LH (i.e., mixed, with the fractions endogenously determined). Within the HH and LL groups everyone will get matched (assuming even numbers). Within the HL group the outcomes depends on the fraction of the two types, but we assume that the maximum possible number of matches are formed.

In this model there are spillovers from the decisions of the Hs on the decisions of the Ls. If Hs decide to stop matching with the Ls, then Ls might be forced to change their matching habits. However for reasons that will become clear, this model does not deliver the desired patterns.

In non-MF villages we have assumed that the payoffs for Hs and Ls are identical and therefore there are many possible equilibria. However, in all equilibria the shares of H and L types in the LH group must be the same.

In MF villages, observe that

$$\begin{aligned}\Delta v_{HL} - \Delta v_{HH} &= \alpha_H \Delta\beta_H b + \beta_H \Delta\alpha_H r - (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r+b) \\ &= -\alpha_H \beta_H \left[\frac{\Delta\beta_H}{\beta_H} r + \frac{\Delta\alpha_H}{\alpha_H} b \right].\end{aligned}$$

This leaves us with two possibilities. Either $\frac{\Delta\beta_H}{\beta_H} r + \frac{\Delta\alpha_H}{\alpha_H} b > 0$ or not. Assume the expression is positive. Since we started from a situation where $v_{HL} = v_{HH} = v_{LL}$, the condition implies that in MF villages $v_{HH} > v_{HL}$. Therefore all Hs will chose the HH option. Paradoxically the same condition also tells us that $\Delta v_{LH} > 0$, so in MF villages $v_{LL} < v_{LH}$. In other words, an L will prefer to be matched with an H. However, the probability of being matched with an H

is zero for an L, since all Hs will choose the HH option. Therefore all Ls will choose the LL option.

Or second, $\frac{\Delta\beta_H}{\beta_H}r + \frac{\Delta\alpha_H}{\alpha_H}b < 0$. In this case Hs will want to match with Ls but not the other way around. Therefore once again we will see full homophily. The fraction of both the HH and LL populations will go up and that of HL will go down in both cases. However in both cases the value of HH links has gone down ($\Delta v_{HH} < 0$), while that of LL links is unchanged. Therefore the fraction of HH links actually formed may go up or down. The fraction of LL links should however go up and therefore on aggregate, the LH population turns into HHs and LLs in MF villages. Randomly formed LL pairs out of this population have the same probability of turning into an actual link as randomly formed LH pairs, but randomly formed HH pairs have lower chance of turning into an actual link. The total number of realized links should therefore be lower in MF villages.

This example is extreme but it captures a robust intuition. If microfinance makes Ls want to pair with Hs rather than with Ls, it also makes Hs want to pair with Hs, and vice versa, which is why there are no LH pairs in MF villages.

Claim **4.** *If new links are formed by directed matching, the introduction of microfinance should*

(1) *either reduce or increase new HH links,*
(2) *reduce new LH links,*
(3) *increase new LL links,*
(4) *and the total number of new links should be less than in non-microfinance villages.*

We can see from the result that the predictions of directed search are inconsistent with the data, namely because the effect on LLs should be positive rather than negative, whereas the number of LH links would go down.

D.5. **Adding More General Dependencies to Our Model.** We now describe an extension of our model to include dependencies in link presence. We describe a variation on the subgraph formation model of Chandrasekhar and Jackson (2018).

Let G be some set of potential subgraphs on n nodes. For instance, instead of just a list of all possible links, it could also include triangles, or various other cliques, stars, and so forth. Let $g \in G$ denote a particular subgraph. In the body we have considered G to be only links, while in the extension, G includes links and triangles. With slight abuse of notation, let $i \in g$ for some $g \in G$ denote that i is one of the nodes that has links in g. Let $v_i(g)$ denote the utility of i if g forms. The total utility that i obtains is the sum over all subgraphs that i is a part of - so rather than just a network, the resulting object is a multigraph.

We let m_g denote a relative frequency adjustment for the type of subgraph in question, as some may be more or less likely to form as a function of the efforts, as in the body, given by

e_i. The individual once again picks a socializing effort e_i and, conditional on socializing, meets and interacts with others, forming various subgraphs via mutual consent.

The probability that some g forms if it is not present is then

$$m_g \times_{i \in g} e_i(1 - F(v_i(g)))$$

which is the product of the socialization efforts and the probability that each i involved in g finds it valuable to form g.

The probability that a subgraph is maintained if it is already present is[50]

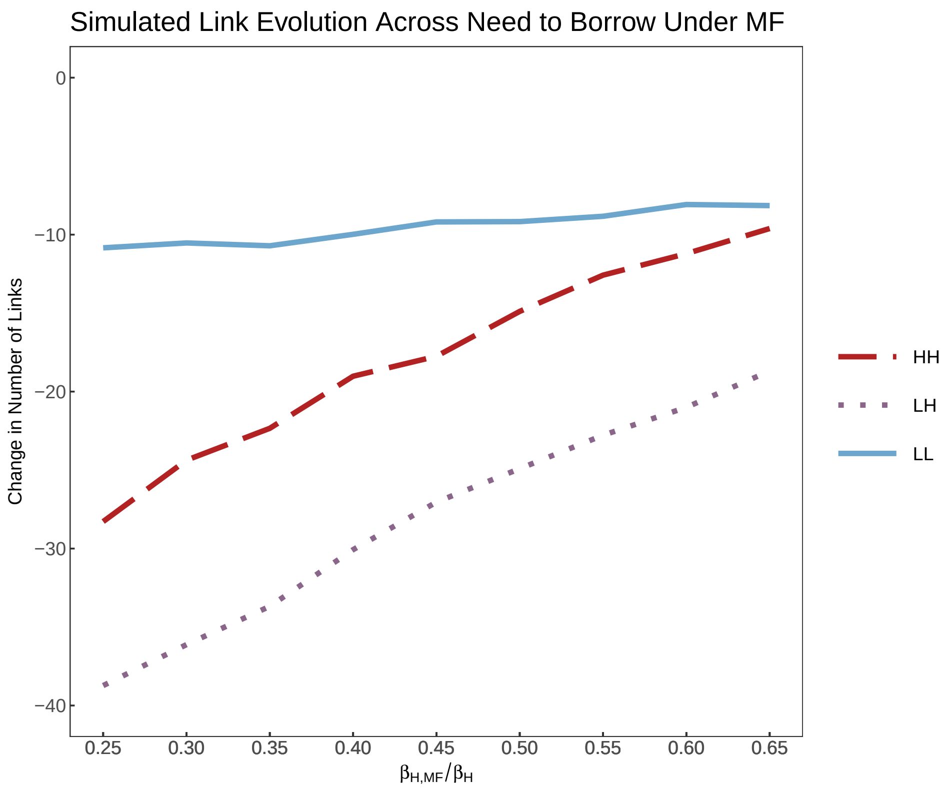
$$\times_{i \in g} e_i.$$

Letting $E^+[v_i(g)]$ denote the expected utility that i gets from subgraph g conditional on finding it worthwhile to form and $\mathcal{G}^t$ denote the set of subgraphs present at the beginning of time t, then the expected utility that i gets from effort e_i is

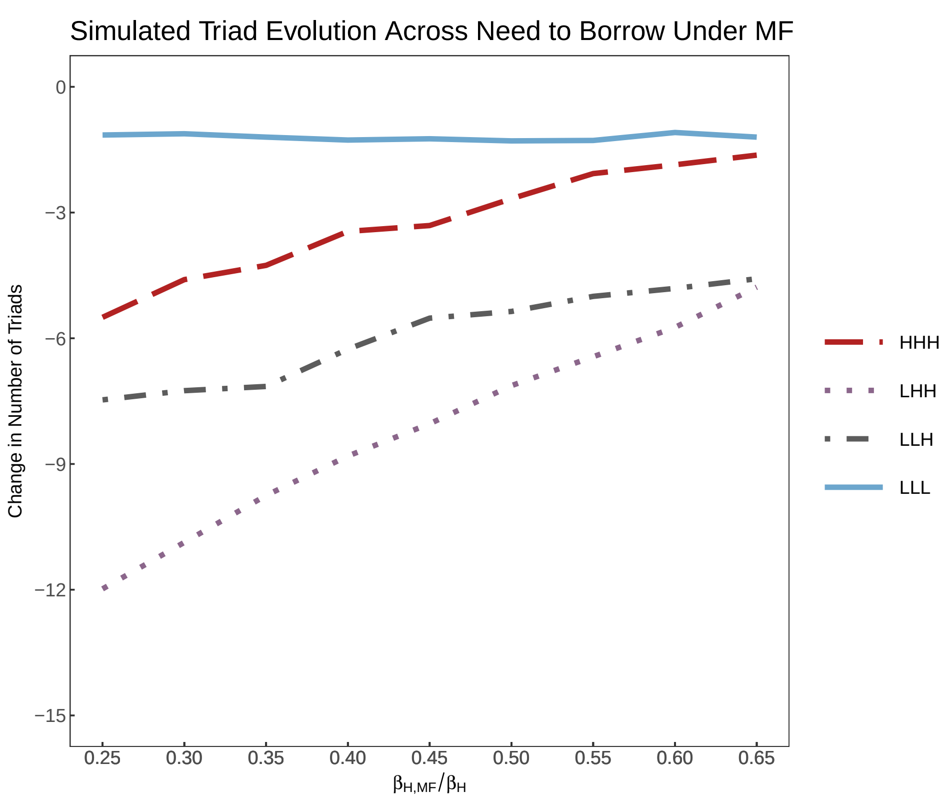
$$\begin{aligned} V_i\left(e_i\right) = u_\theta e_\theta - \frac{1}{2} c_\theta e_\theta^2 + \sum_{g \in \mathcal{G}^t} \mathrm{E}^+[v_i(g)] \times_{j \in g} e_j \\ + \sum_{g \notin \mathcal{G}^t} \mathrm{E}^+[v_i(g)] m_g \times_{j \in g} e_j(1 - F(v_j(g))). \end{aligned}$$

The model then functions just as the model described in the text, simply with these augmented preferences over richer collections of subgraphs.

[50]One could adjust the relative impact of effort for maintaining a subgraph to be some other function than simply the product, depending on the context.



(A) Evolution of HH, LH, and LL



(B) Evolution of HHH, LHH, LLH, and LLL

Figure D.1. Supported Links Model

Online Appendix: Not for Publication

Appendix E. Hyderabad Network Elicitation

E.1. **Survey Questions.**

E.1.1. *Direct Link Elicitation.* We first ask the following set of network questions

(1) *Financial* relationships
   (a) If your gas cylinder, kerosene or any other cooking fuel runs out while cooking and you don't have it readily available at home, who would you go to in this neighborhood to borrow some and who would come to you in a similar situation?
   (b) If you need 50 or 100 Rupees because you're falling short for some payment, who in this neighborhood would you borrow this money from and who from this basti would come to you in a similar situation?
   (c) If you had visitors and needed some milk or sugar to make tea but the shop is closed, who in this neighborhood would you borrow it from and who would come to you in a similar situation?
(2) If you needed advice on financial matters, for example, opening a savings account, buying gold, taking a loan, buying insurance, making investments, etc. who in this neighborhood would you go to and who would come to you for similar advice?
(3) *Information* relationships (non-finance)
   (a) If you needed advice on which school/college to put your children in, who in this neighborhood would you go to and who would come to you for similar advice?
   (b) If you had to move to another house in this neighborhood, who would you ask for help to find a house and who would come to you for help to find a house?
   (c) If your child or another member of your family falls sick, who in this neighborhood would you go to for advice and who would come to you for similar advice?
(4) *Social* relationships
   (a) Who would come or send their children to your house to watch television and whose house would you or your children go to for the same purpose?

While these questions resemble those in a full network elicitation, there are several key differences. First, we only interview a subsample of the neighborhood. Second, we do not have a census enumeration of the full neighborhood, so consequently, third, we do not attempt to match survey responses to form an adjacency matrix.

E.1.2. *ARD Questions.* We collected Aggregated Relational Data (ARD) using the following questions:

How many other households do you know in your neighborhood ...

(1) where a woman has ever given birth to twins?

(2) where there is a permanent government employee?
(3) where there are 5 or more children?
(4) where anye child has studied past 10th standard?
(5) where any adult has had typhoid, malaria, or cholera in the past six months?
(6) where any adult has been arrested by the police?
(7) where at least one woman has had a second marriage?
(8) where at least one man currently has more than one wife?
(9) where at least one member has migrated abroad for work?

Each respondent was also asked whether her household posessed each of these traits.

E.2. **ARD Algorithm.** We adapt the ARD algorithm from Breza et al. (2020b). Here we provide an overview of the method. Suppose that a researcher is interested in studying networks in a set of distinct communities. A network with n households is given by $\mathbf{g}$, which is a collection of links ij where $g_{ij} = 1$ if and only if households i and j are linked and $g_{ij} = 0$ otherwise. Our goal is to estimate characteristics of g, such as the probability that arbitrary pairs or triples of nodes are linked.

I. **Conduct ARD survey:** Sample a share ψ (e.g., 30 percent) of households. Have each enumerate a list of their network links. Note that this gives a direct estimate of the respondent's degree. Ask 5-9 ARD questions, such as

"How many households among your network list do you know where any adult has had typhoid, malaria, or cholera in the past six months?"

The ARD response for a household i is

$$y_{ik} = \sum_j g_{ij} \cdot \mathbf{1}\{j \text{ has had one of those diseases in past 6 mo.}\}$$

where trait k denotes the disease question. This adds up all links that have had the diseases over the last six months. Ask whether the respondent household themselves has each ARD trait k as well to generate population estimates for the prevalence of each trait.

II. **Estimate network formation model with ARD:** Use the information from the ARD survey and the trait prevalences to estimate the parameters of a network formation model. In this model, the probability that two households i and j are linked depends on household fixed effects (ν_i) and distance in a latent space (latent locations z_i) with

$$\mathrm{P}(g_{ij} = 1 | \nu_i, \nu_j, \zeta, z_i, z_j) \propto \exp(\nu_i + \nu_j + \zeta \cdot \mathrm{distance}(z_i, z_j)).$$

We use a Bayesian framework to estimate the model parameters and latent space locations. We assume a latent space on $\mathcal{S}^3$, the surface of a sphere. Priors for latent positions of each individual and trait group follow a von Mises-Fisher distribution:

$$z_i | v_z, \eta_z \sim \mathcal{M}(v_z, 0)$$

$$z_{j\in G_k}|\upsilon_k, \eta_k \sim \mathcal{M}(\upsilon_k, \eta_k)$$

The centers $\{\upsilon_z, \upsilon_k\}_{z,k}$ and concentration parameters $\{\eta_k\}_k$ of the prior distributions also need to be estimated.

For respondent i and subpopulation G_k, recall that we observe:

$$y_{ik} = \sum_{j\in G_k} g_{ij}.$$

Conditional on latent positions, $(z_i, z_{j\in G_k})$, and household fixed effects $\{\nu_i\}_i$, y_{ik} approximately follows a Poisson distribution when the number of individuals in the subpopulation n_k is large. The Poisson parameter is:

$$\lambda_{ik} = \sum_{j\in G_k} \mathrm{P}(g_{ij} = 1|\nu_i, \nu_{j\in G_k}, \zeta, z_i, z_{j\in G_k})$$

The expected ARD response by i for category k can be expressed as

$$\lambda_{ik} = \mathrm{E}[y_{ik}] = d_i b_k \left(\frac{C_{p+1}(\zeta)C_{p+1}(\eta_k)}{C_{p+1}(0)C_{p+1}\sqrt{\zeta^2 + \eta_k^2 + 2\zeta\eta_k \cos(\theta_{(z_i,\upsilon_k)})}} \right),$$

where d_i is the respondent expected degree, $b_k = n_k/n$ the share of nodes in group k, $C_{p+1}(\cdot)$ is the normalizing constant of the von Mises-Fisher distribution, and $\theta_{(z_i,\upsilon_k)}$ is the angle between the two vectors. Note that the resulting likelihood relies entirely on ARD:

$$y_{ik}|d_i, b_k, \zeta, \eta_k, \theta_{(z_i,\upsilon_k)} \sim \text{Poisson}\left(\lambda_{ik}\right).$$

Finally, to estimate the household fixed effects $\{\nu_i, \nu_j\}$, note that

$$d_i = n \exp(\nu_i)\mathrm{E}[\exp(\nu_j)] \left(\frac{C_{p+1}(0)}{C_{p+1}(\zeta)} \right).$$

Letting $\boldsymbol{\theta}$ be a shorthand for all parameters, we can estimate the posterior

$$\begin{aligned}\boldsymbol{\theta}|y_{ik} \propto & \prod_{k=1}^{K}\prod_{i=1}^{n} \exp(-\lambda_{ik})\lambda_{ik}^{y_{ik}} \prod_{i=1}^{n} \text{Normal}(\log(d_i)|\mu_d, \sigma_d^2) \\ & \times \prod_{k=1}^{K} \text{Normal}(\log(b_k)|\mu_b, \sigma_b^2) \prod_{k=1}^{K} \text{Normal}(\log(\eta_k)|\mu_{\eta_k}, \sigma_{\eta_k}^2)\text{Gamma}(\zeta|\gamma_\zeta, \psi_\zeta).\end{aligned}$$

Given this posterior distribution, the probability of any network $\mathbf{g}$ being drawn is fully computed.

III. **Compute network statistics of interest:** Use the estimated probability model (using ζ, fixed effects ν_i and latent locations z_i) to compute $\mathrm{E}[S(\mathbf{g})|\mathbf{Y}]$, where $S(\mathbf{g})$ is, for example, the probability of a link between any arbitrary pair (or triple) of nodes. The data and replication code is freely available (Breza et al., 2020a).

Appendix F. Classification and ARD Robustness

F.1. **Classification of H and L Types.** In our main specifications, we follow the description in Appendix B when classifying households into H and L types and use the resulting type indicators directly in our analysis. In what follows, we re-estimate our main regression tables accounting for the fact that the classification of H types is done with noise.

Specifically, we conduct a bootstrap procedure where we take many draws of the training data and re-run the random forest classification exercise for each draw. Within a draw, we re-estimate our main regressions of interest in a block bootstrap, sampling from villages or neighborhoods with replacement. In practice, we take 100 separate draws from the training data, with five block bootstrap draws each. This gives $100 \times 5 = 500$ total parameter estimates for each regression coefficient of interest. We report the median value across all 500 draws and construct confidence intervals and p-values from the distribution of estimated parameters.

We find that the results look quite similar to our main tables.

Table F.1. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF
Microfinance	-0.060	-0.020
	(-0.097,-0.028)	(-0.035,-0.007)
	[0.000]	[0.004]
Microfinance $\times$ LH	0.011	0.001
	(-0.019,0.045)	(-0.005,0.009)
	[0.536]	[0.724]
Microfinance $\times$ HH	0.039	0.002
	(-0.008,0.087)	(-0.012,0.017)
	[0.160]	[0.748]
LH	-0.025	0.002
	(-0.061,-0.004)	(-0.005,0.008)
	[0.052]	[0.596]
HH	0.009	0.024
	(-0.038,0.049)	(0.013,0.038)
	[0.820]	[0.000]
Observations	57,376	846,561
Linked Pre-MF	Yes	No
Controls	No	No
Depvar Mean	0.441	0.0636

Notes: 90% CI are block-bootstrapped at the village level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

TABLE F.2. Link Evolution, Hyderabad

	(1) Prob. Linked
Microfinance	-0.006
	(-0.01,-0.002)
	[0.016]
Microfinance × LH	0.001
	(-0.008,0.006)
	[0.808]
Microfinance × HH	-0.005
	(-0.024,0.009)
	[0.568]
LH	0.002
	(-0.003,0.011)
	[0.608]
HH	0.011
	(-0.002,0.031)
	[0.16]
Observations	141,996
Controls	No
Depvar Mean	0.0255

Notes: 90% CI are block-bootstrapped at the neighborhood level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. The controls are selected by double post lasso among all the variables that are used for its random forest classification, and includes several household and village level characteristics.

TABLE F.3. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Any link in triangle survived Post-MF
Microfinance	-0.078	-0.080
	(-0.141,-0.033)	(-0.124,-0.041)
	[0.000]	[0.000]
Microfinance × LLH	0.029	0.031
	(-0.013,0.076)	(-0.006,0.064)
	[0.240]	[0.180]
Microfinance × LHH	0.061	0.057
	(0.008,0.121)	(0.006,0.104)
	[0.064]	[0.076]
Microfinance × HHH	0.098	0.082
	(0.018,0.180)	(0.021,0.143)
	[0.056]	[0.024]
LLH	-0.029	-0.028
	(-0.079,0.005)	(-0.058,0.002)
	[0.132]	[0.128]
LHH	-0.037	-0.025
	(-0.105,0.005)	(-0.074,0.009)
	[0.136]	[0.24]
HHH	-0.024	-0.009
	(-0.091,0.051)	(-0.06,0.04)
	[0.480]	[0.792]
Observations	53,233	53,233
Linked Pre-MF	Yes	Yes
Controls	No	No
Depvar Mean	0.197	0.808

Notes: 90% CI are block-bootstrapped at the village level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

TABLE F.4. Triples Evolution, Hyderabad

All variables x 1000	Full Triangle Linked
	(1)
Microfinance	-0.018
	(-0.04,-0.003)
	[0.04]
Microfinance × LLH	0.002
	(-0.039,0.023)
	[0.852]
Microfinance × LHH	-0.022
	(-0.161,0.039)
	[0.668]
Microfinance × HHH	-0.079
	(-0.328,0.043)
	[0.316]
LLH	0.005
	(-0.015,0.047)
	[0.852]
LHH	0.043
	(-0.013,0.182)
	[0.304]
HHH	0.113
	(-0.001,0.393)
	[0.108]
Observations	3,341,006
Controls	No
Depvar Mean	3.53e-02

Notes: 90% CI are block-bootstrapped at the neighborhood level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. The controls are selected by double post lasso among all the variables that are used for its random forest classification, and includes several household and village level characteristics.

Table F.5. Borrowing patterns

Panel A: Borrowing Patterns, Karnataka

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance × Post	543	-530	-819	550	268
	(271,777)	(-1118,-26)	(-1453,-259)	(-905,1973)	(-841,1562)
	[0.000]	[0.088]	[0.024]	[0.480]	[0.660]
Microfinance × Post × H	1656	204	-25	-2096	-548
	(1068,2292)	(-306,766)	(-600,704)	(-3883,93)	(-2566,1465)
	[0.000]	[0.496]	[1.000]	[0.112]	[0.732]
Microfinance × H	80.6	-33.4	11	794	481
	(-55,213)	(-164,71)	(-476,478)	(-414,1749)	(-1098,2023)
	[0.352]	[0.584]	[0.948]	[0.288]	[0.788]
Post × H	219	-364	181	1619	-68
	(56,448)	(-821,28)	(-433,705)	(46,3074)	(-1337,1133)
	[0.040]	[0.148]	[0.688]	[0.084]	[0.972]
Observations	28,062	27,194	28,062	28,062	28,062
Depvar Mean	596.976	860.228	1863.324	2667.56	1656.881

Panel B: Borrowing Patterns, Hyderabad

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance	-79	286	-2165	-2001	149
	(-849,565)	(-2330,3157)	(-4485,840)	(-6596,2037)	(-1908,2205)
	[0.780]	[0.868]	[0.240]	[0.396]	[0.964]
Microfinance × H	7688	-1210	-320	1891	544
	(5826,9408)	(-4783,2071)	(-3248,2807)	(-3424,8496)	(-1859,3685)
	[0.000]	[0.516]	[0.784]	[0.520]	[0.652]
H	431	-1272	1563	1168	7
	(-266,1246)	(-4349,2048)	(-1802,3813)	(-4516,5584)	(-2222,1784)
	[0.268]	[0.656]	[0.468]	[0.952]	[1.000]
Observations	6,811	6,863	6,863	6,863	6,863
Depvar Mean	3107.86	7895.05	6935.66	18805.06	2620.97

Notes: These tables present the effect of microfinance access on the loan amounts borrowed from various sources. Panel A pertains to Karnataka (outcomes are winsorized to the 2.5% level) and tracks loan amounts from microfinance institutions, friends, family, banks and moneylenders. All of its columns control for surveyed in wave 1 fixed effects. The average rate of inflation over the period between waves in Karnataka was 8% meaning a total of 65%. Panel B pertains to Hyderabad (outcomes are winsorized to the 2.5% level). Outcomes are measured in the first survey wave (2007-08). Here all specifications include demographic household and village controls (the same ones used in random forest classification of H vs L) subject to double-post LASSO. 90% CI are block-bootstrapped at the village/neighborhood level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. MFI: Microfinance Institution;SHG: Self-Help Group.

Table F.6. Risk sharing, Hyderabad

	(1) Expenditures: Non-Food	(2) Expenditures: Total
Microfinance × Income	0.061	0.048
	(0.002,0.116)	(-0.022,0.110)
	[0.084]	[0.276]
Microfinance × Income × H	-0.024	-0.047
	(-0.126,0.086)	(-0.171,0.106)
	[0.740]	[0.696]
Household Income per capita	0.061	0.118
	(0.030,0.105)	(0.079,0.171)
	[0.004]	[0.000]
Household Income per capita × H	0.004	0.043
	(-0.065,0.069)	(-0.060,0.143)
	[0.976]	[0.648]
Observations	10,502	10,590
Depvar Mean	1193	2040
Income Mean	1440	1437

Notes: Income is total household, monthly per capita earnings from employment or business activities, excluding private and government transfers. Dependent variable is monthly per capita household expenditure. In col. 1, expenditure excludes food and in col. 2, we present total expenditure. Data is from the first (2007-08) and third (2012) waves of the Hyderabad survey. Regression includes controls for household fixed effects and wave-by-neighborhood-by-type fixed effects. Additional controls are selected by double post lasso from the set of variables used in the prediction exercise, interacted with type. 90% CI are block-bootstrapped at the neighborhood level across 5 bootstraps times 100 H classifications (trained on different samples). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets.

F.2. **Aggregated Relational Data.** Recall that in the Hyderabad dataset, we do not observe links and triples directly, but rather estimate the linking probabilities using ARD. In what follows, we outline a procedure to take in to account the estimation error from using ARD.

We build from the fact that our ARD procedure produces a full posterior distribution of bilateral linking probabilities. Moreover, once we observe bilalteral linking probabilities, it is straightforward to calculate the probability of triangles for each potential triple of nodes. We again proceed using a two step bootstrap procedure. In the first step, we randomly draw the linking probabilities from the posterior distribution. In the second step, we use those linking probabilities in regressions, but additionally conduct a block bootstrap, resampling neighborhoods with replacement. In practice, we take 450 draws from the posterior distribution and 5 draws each in the block boostratp step. Thus, the procedure generates $450 \times 5 = 2250$ parameter values for each coefficient of interest. As before, we report the median value across all 2250 draws and construct confidence intervals and p-values from the distribution of estimated parameters. We once again find that the results look quite similar to our main specification.

Table F.7. Hyderabad: Characteristics of H versus L (Network Variables)

	(1) Exp. Degree	(2) Exp. Links to L	(3) Exp. Links to H	(4) Exp. Eig. Cent.
H	0.182	-0.651	0.834	0.009
	(-0.164,0.541)	(-0.653,0.096)	(0.087,0.835)	(-0.002,0.015)
	[0.403]	[0.248]	[0.032]	[0.235]
Observations	4,523	4,523	4,523	4,523
Depvar Mean	5.813	4.353	1.463	0.075

Notes: 90% CI are block-bootstrapped at the village level across the probabilities to be linked (drawn from the ARD posterior distribution of the parameters). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. The estimates reflect H-vs.L differences for the non-microfinance (control group) sample only.

Table F.8. Hyderabad: Graph-Level Characteristics

	(1) Clustering	(2) Clustering	(3) Closeness	(4) Closeness
Microfinance	-0.003	-0.003	-0.001	-0.001
	(-0.020,0.011)	(-0.020,0.011)	(-0.008,0.004)	(-0.007,0.004)
	[0.691]	[0.717]	[0.766]	[0.840]
Double-Post LASSO	No	Yes	No	Yes
Depvar Mean	0.0352	0.0352	0.00274	0.00274
Non MF Mean	0.0375	0.0375	0.00325	0.00325
Observations	89	89	89	89

Notes: 90% CI are block-bootstrapped at the village level across the probabilities to be linked (drawn from the ARD posterior distribution of the parameters). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. Controls are the same demographic characteristics of households and villages that are used in random forest classification of H vs L.

Table F.9. Link Evolution, Hyderabad

	(1) Prob. Linked
Microfinance	-0.005
	(-0.01,-0.001)
	[0.026]
Microfinance × LH	0.002
	(-0.004,0.007)
	[0.591]
Microfinance × HH	-0.011
	(-0.025,0.004)
	[0.228]
LH	0.002
	(-0.003,0.008)
	[0.506]
HH	0.018
	(0.005,0.032)
	[0.016]
Observations	141,996
Controls	No
Depvar Mean	0.0255
LL, Non MF Mean	0.0268

Notes: 90% CI are block-bootstrapped at the village level across the probabilities to be linked (drawn from the ARD posterior distribution of the parameters). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. The controls are selected by double post lasso among all the variables that are used for its random forest classification, and includes several household and village level characteristics.

TABLE F.10. Triples Evolution, Hyderabad

All variables x 1000	Full Triangle Linked
	(1)
Microfinance	-0.018
	(-0.044,-0.002)
	[0.062]
Microfinance × LLH	0.01
	(-0.012,0.035)
	[0.382]
Microfinance × LHH	-0.027
	(-0.136,0.036)
	[0.465]
Microfinance × HHH	-0.177
	(-0.493,-0.038)
	[0.036]
LLH	0
	(-0.019,0.023)
	[0.967]
LHH	0.055
	(0.005,0.163)
	[0.068]
HHH	0.217
	(0.1,0.548)
	[0.007]
Observations	3,341,006
Controls	No

Notes: 90% CI are block-bootstrapped at the village level across the probabilities to be linked (drawn from the ARD posterior distribution of the parameters). These 90% CI are reported in parentheses. Estimated p-values are reported in brackets. The controls are all the variables that are used for its random forest classification, and includes several household and village level characteristics.

Appendix G. Supplemental Borrowing Results

Here, we look in more detail at the increases and declines in informal borrowing amounts, paying attention to the number of Hs the household is linked to at baseline and the degree at baseline. This exercise is only possible in the Karnataka network panel data. This allows us, for example, to concentrate on Ls with no links to Hs and look at the pure externality effect. We estimate

$$\begin{aligned} y_{ivt} = \alpha &+ \beta_1 \mathrm{MF}_v \times \text{No. of } H \text{ links}_{iv} \times \mathrm{Post}_t \\ &+ \gamma_1 \mathrm{MF}_v \times \mathrm{Post}_t + \gamma_2 \text{No. of } H \text{ links}_{iv} \times \mathrm{Post}_t + \gamma_3 \mathrm{MF}_v \times \text{No. of } H \text{ links}_{iv} \\ &+ \delta_1 \mathrm{MF}_v + \delta_2 \text{No. of } H \text{ links}_{iv} + \delta_3 \mathrm{Post}_t \\ &+ \eta_1 \mathrm{MF}_v \times \text{No. of links}_{iv} \times \mathrm{Post}_t + \eta_2 \mathrm{MF}_v \times \text{No. of links}_{iv} + \eta_3 \text{No. of links}_{iv} + \epsilon_{ivt}, \end{aligned}$$

where again y_{ivt} is the amount borrowed from the stated source (MFI, friends, self-help group, family, moneylenders), No. of links is the degree in Wave 1, and No. of H links is the baseline number of H links in Wave 1.

We are particularly interested in $\gamma_1 + \eta_1 \cdot d$, the differential effect of being in a microfinance village in the second period, without any H-type of links at baseline and just d of L-type links at baseline. This reflects the pure externalities among people who had nothing to do with microcredit whatsoever as they have no H links. For example, if the household had only one L link and no H links, then $\gamma_1 + \eta_1$ is the differential effect of being in the second period in a village exposed to microfinance.

Table G.1 presents the results. In Panel A we look at H respondents and in Panel B we look at L respondents. Loans from friends go down for both types (not significantly for H) when exposed to microfinance. Strikingly the fall is greater for L households. For an L household with a single L friend and no H friends, the effect is $\gamma_1 + \eta_1$, which corresponds to a decline of Rs. 2572.3 ($p = 0.015$) more than a comparable H household (a decline of Rs. 1194, not significantly different from zero). We can statistically reject that the decline of Rs. 2572 for L is smaller than the Rs. 1194 ($p = 0.094$) for H.

We also find that Hs engage in complementary borrowing from moneylenders. An H household with one L link borrows Rs. 2379 more from moneylenders when exposed to microfinance ($p = 0.04$).

Table G.1. Borrowing patterns by link composition, Karnataka

Panel A: H Nodes

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance × No. of Hs × Post	105.260	−49.522	−5.536	−99.058	−116.978
	(28.573)	(22.830)	(87.826)	(112.745)	(151.521)
	[0.000]	[0.031]	[0.950]	[0.380]	[0.441]
MF × Post	2,481.568	−1,193.977	187.085	−663.858	1,804.001
	(342.864)	(732.601)	(545.908)	(958.570)	(1,586.261)
	[0.000]	[0.104]	[0.732]	[0.489]	[0.256]
Depvar Mean	923.174	825.701	2250.665	2971.487	1542.522
Observations	10,183	9,949	10,183	10,183	10,183

Panel B: L Nodes

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance × No. of Hs × Post	73.072	−97.247	−83.646	96.473	10.941
	(18.768)	(34.840)	(69.934)	(132.281)	(109.634)
	[0.000]	[0.006]	[0.232]	[0.466]	[0.921]
MF × Post	572.440	−1,294.876	−751.320	409.621	−1,660.386
	(134.751)	(559.610)	(646.798)	(958.405)	(1,051.831)
	[0.000]	[0.021]	[0.246]	[0.670]	[0.115]
Depvar Mean	411.136	880.15	1642.649	2494.409	1722.033
Observations	17,879	17,245	17,879	17,879	17,879

Notes: This table presents the effect of microfinance access, no. of H neighbours and their interactions on the loan amounts borrowed from microfinance institutions, friends, family, banks and moneylenders. Panel A conditions on H nodes, and Panel B conditions on L nodes. All columns control for surveyed in wave 1 fixed effects. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Appendix H. Results for Unsupported Links and Dependencies in Trangles

H.1. **Results for Unsupported LL Links.** In the Karnataka data, we can further examine the evolution of pre-existing LL links separately for supported and unsupported links. Here, we restrict the sample to all LL links that exist at baseline (Wave 1) and regress whether a link $g_{ij,2}$ exists in Wave 2 on whether the village had microfinance, whether the households had links in common, and interactions. Specifically,

$$\begin{aligned} g_{ij,v,2} &= \alpha + \beta MF_v + \beta_{FIC}\text{No. Friends in Common}_{ij,v} \\ &+ \beta_{FIC,MF}\text{No. of Friends in Common}_{ij,v} \times MF_v \\ &+ \beta_H \frac{\text{No. of High FIC}}{\text{No. of FIC}}_{ij,v} + \beta_{H,MF}\frac{\text{No. of High FIC}}{\text{No. of FIC}}_{ij,v} \times MF_v + \epsilon_{ij,v,2}, \end{aligned}$$

(where FIC is "Friends in Common")

Table H.1 presents the result.

Table H.1. Evolution of LL links as a function of support from H links (Karnataka)

	(1) Linked Post-MF
Num High Friends in Common / Num Friends in Common × MF	0.041
	(0.033)
	[0.221]
Num High Friends in Common / Num Friends in Common	-0.054
	(0.028)
	[0.057]
Microfinance	-0.047
	(0.027)
	[0.086]
Num Friends in Common × MF	-0.014
	(0.008)
	[0.064]
Num Friends in Common	0.018
	(0.007)
	[0.007]
Linked Pre-MF	Yes
LL links only	Yes
Depvar Mean	0.436
Observations	18,712

Notes: Regression includes fixed effects for number of friends in common and interaction of these dummies with MF. Number of High Friends in Common has a mean of 0.49 with a standard deviation of 1.03. Number of High Friends in Common / Number of Friends in Common has a mean of 0.61 with a standard deviation of 0.42. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

H.2. **Dependencies in Maintenance of Triangles.** We assess the relevance of link interdependencies in our setting in the Karnataka panel data in non-microfinance villages. We consider triples of L nodes that are all connected in Wave 1 and ask whether those triangles are more likely to survive to Wave 2 than what we would expect under link independence. In Table H.1, we show that when LL links are unsupported (i.e., there is no third node with whom both members in the pair have a relationship), the probability of survival is 0.436. Thus, under independence, the survival probability for a full LLL trangle is $(0.436)^3 = 0.083$. However, we find that 27.75% of LLL triples observed in Wave 1 survive to Wave 2 (Table 7), which is substantially higher than what we would expect under independence. This suggests a potentially important role for local externalities and link dependencies.

Appendix I. Building Social Capital among MF Takers cannot Explain Results

In Table I.1 we repeat our main regression of whether a link exists in Wave 2 as a function of microfinance exposure and interactions with household type. In columns 2 and 4 we include indicators for whether at least one of the households involved joined microfinance, so the main effects are for households not involved in microfinance whatsoever. (This is clearly not to be causally interpreted, but merely illustrative.)

From this we see that our results are essentially unchanged. That is, for household pairs of type HH, HL, or LL, when a link exists in Wave 1, the greatest relative declines in the probability of the link surviving in MF villages versus non-MF villages come from LL and HL, rather than HH. The differential effects of having the households (typically Hs) join microfinance could not have been driving our main result (the interactions are insignificant and have small point estimates). A similar phenomenon holds in column 4. Thus, we find that for the vast majority of pairs, which are not at all involved in microfinance, in microfinance villages they experience a relative decline in probability of being linked in the second period and that the decline is larger for LL than for HH pairs.

Then in Table I.2 we regress whether a link exists in Wave 2 further interacting by whether one or both of the households involved joined microfinance. We can see again that our main results (for those who have no parties joining microfinance) are unchanged, demonstrating that our results are not driven by new links among microfinance members. However, it is interesting to note that HH pairs that both enroll in microfinance and that are previously unlinked are considerably more likely (1.5pp relative to a mean of 6.4%) to form a new link, consistent with Feigenberg et al. (2013). Of course, the main effect of having microfinance for this pair is a 2.2pp decline in the probability of forming the new link to begin with, so this means that on net there is no effect: the new relationships forged by meeting others in microfinance centers serve only to offset the greater decline in social capital overall.

Taken together, we see that (a) even looking at parties that never joined microfinance, LL types experience greater social capital losses than HH and (b), while HHs involved in microfinance are able to stave off some of the loss in linking rates in MF villages because microfinance takers wind up forming some links to each other, they are not nearly numerous enough to drive our main results (noting that 86% of pairs households in microfinance villages involve households that did not take-up).

Table I.1. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance	−0.058	−0.057	−0.023	−0.022
	(0.018)	(0.018)	(0.008)	(0.008)
	[0.002]	[0.002]	[0.006]	[0.007]
Microfinance × HH	0.039	−0.016	0.009	0.005
	(0.022)	(0.027)	(0.007)	(0.009)
	[0.085]	[0.534]	[0.206]	[0.560]
Microfinance × LH	0.009	0.007	0.007	0.005
	(0.015)	(0.016)	(0.004)	(0.004)
	[0.573]	[0.675]	[0.120]	[0.257]
LH	−0.025	−0.025	−0.002	−0.002
	(0.012)	(0.012)	(0.004)	(0.004)
	[0.036]	[0.036]	[0.566]	[0.566]
HH	0.008	0.008	0.021	0.021
	(0.017)	(0.017)	(0.006)	(0.006)
	[0.622]	[0.623]	[0.001]	[0.001]
Microfinance x HH x At least one in MF		0.089		0.010
		(0.025)		(0.007)
		[0.001]		[0.178]
Microfinance x LH x At least one in MF		0.018		0.008
		(0.023)		(0.003)
		[0.422]		[0.016]
Microfinance x At least one in MF		−0.017		−0.005
		(0.022)		(0.003)
		[0.441]		[0.070]
Observations	57377	57377	846562	846562
Linked Pre-MF	Yes	Yes	No	No
Depvar Mean	0.441	0.441	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.0753	0.0753
MF + MF x $LH = 0$ p-val	0.364	0.003	0.101	0.074
MF + MF x $HH = 0$ p-val	0.014	0.014	0.015	0.01
MF + LHxMF = MF + HHxMF p-val	0.136	0.332	0.642	0.97

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Table I.2. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF
Microfinance	−0.057	−0.022
	(0.018)	(0.008)
	[0.002]	[0.007]
Microfinance × HH	−0.016	0.005
	(0.027)	(0.009)
	[0.534]	[0.560]
Microfinance × LH	0.007	0.005
	(0.016)	(0.004)
	[0.675]	[0.257]
LH	−0.025	−0.002
	(0.012)	(0.004)
	[0.036]	[0.566]
HH	0.008	0.021
	(0.017)	(0.006)
	[0.623]	[0.001]
One takes MF	−0.013	−0.005
	(0.022)	(0.003)
	[0.544]	[0.061]
Both take MF	−0.101	−0.003
	(0.050)	(0.014)
	[0.043]	[0.818]
MF x LH x One takes MF	0.009	0.007
	(0.024)	(0.003)
	[0.701]	[0.026]
MF x HH x One takes MF	0.045	−0.001
	(0.028)	(0.007)
	[0.106]	[0.843]
MF x LH x Both take MF	0.212	0.025
	(0.044)	(0.013)
	[0.000]	[0.052]
MF x HH x Both take MF	0.222	0.024
	(0.053)	(0.016)
	[0.000]	[0.129]
Observations	57377	846562
Linked Pre-MF	Yes	No
Depvar Mean	.441	.0636
LL, Non-MF Mean	.483	.075
MF + MF × $HH = 0$ p-val	.617	.007
MF + MF × $LH = 0$ p-val	.001	.001
MF × HH = MF × LH p-val	.006	.931

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Appendix J. Alternate Logistic Classification

J.1. **Karnataka Exhibits.**

Table J.1. Characteristics of H versus L, Karnataka

Panel A: Karnataka - Demographics and Amenities variables

	(1) GMOBC	(2) Latrine	(3) Num. Rooms	(4) Num. Beds	(5) Thatched Roof	(6) RCC Roof
H	−0.482	−0.278	−0.602	−0.635	0.027	−0.130
	(0.014)	(0.016)	(0.033)	(0.053)	(0.004)	(0.010)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Depvar Mean	0.7	0.261	2.36	0.84	0.0235	0.117
Observations	14,904	14,904	14,904	14,904	14,904	14,904

Panel B: Karnataka - Network variables

	(1) Degree	(2) Links to L	(3) Links to H	(4) Eig. Cent.
H	3.125	−0.160	3.119	0.019
	(0.194)	(0.164)	(0.149)	(0.002)
	[0.000]	[0.000]	[0.284]	[0.000]
Depvar Mean	8.97	4.05	3.65	0.0524
Observations	14,904	14,904	14,904	14,904

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Panels A and B pertains to Karnataka, based on Wave 1 data only. GMOBC = A dummy for whether the household consists of general caste or backwards caste, so the omitted categories are scheduled caste and scheduled tribes. General and OBC are considered upper caste.

TABLE J.2. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance	−0.062	−0.072	−0.022	−0.019
	(0.020)	(0.020)	(0.009)	(0.007)
	[0.003]	[0.0004]	[0.011]	[0.011]
$LH\times$ MF	0.014	0.011	−0.0002	0.002
	(0.016)	(0.015)	(0.005)	(0.004)
	[0.362]	[0.454]	[0.972]	[0.707]
$HH\times$ MF	0.018	0.014	−0.001	0.001
	(0.023)	(0.020)	(0.007)	(0.007)
	[0.430]	[0.491]	[0.898]	[0.929]
LH	−0.075	−0.061	−0.008	−0.013
	(0.012)	(0.012)	(0.004)	(0.003)
	[0.000]	[0.000]	[0.040]	[0.0001]
HH	−0.021	−0.011	0.021	0.008
	(0.018)	(0.018)	(0.007)	(0.006)
	[0.257]	[0.561]	[0.002]	[0.161]
Observations	57,376	57,376	846,561	846,561
Linked Pre-MF	Yes	Yes	No	No
Controls	No	Yes	No	Yes
Depvar Mean	0.441	0.441	0.0636	0.0636
LL, Non-MF Mean	0.511	0.511	0.0795	0.0795
MF + MF $\times$ $LH = 0$ p-val	0.008	0.001	0.001	0.002
MF + MF $\times$ $HH = 0$ p-val	0.028	0.003	0.007	0.021
MF + $LH\times$ MF = MF + $HH\times$ MF p-val	0.796	0.845	0.879	0.846

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

TABLE J.3. Link Evolution for Info and Financial Links, Karnataka

	(1) Financial Linked Post-MF	(2) Financial Linked Post-MF	(3) Info Linked Post-MF	(4) Info Linked Post-MF
Microfinance	−0.067	−0.012	−0.049	−0.014
	(0.019)	(0.005)	(0.019)	(0.006)
	[0.001]	[0.025]	[0.009]	[0.014]
$LH\times$ MF	0.028	0.0003	0.011	0.001
	(0.016)	(0.003)	(0.015)	(0.003)
	[0.095]	[0.917]	[0.444]	[0.671]
$HH\times$ MF	0.027	−0.003	0.008	−0.002
	(0.021)	(0.004)	(0.022)	(0.004)
	[0.203]	[0.468]	[0.702]	[0.625]
LH	−0.065	−0.006	−0.061	−0.008
	(0.013)	(0.002)	(0.012)	(0.003)
	[0.000]	[0.012]	[0.000]	[0.003]
HH	−0.037	0.015	−0.024	0.013
	(0.016)	(0.004)	(0.018)	(0.004)
	[0.018]	[0.0001]	[0.176]	[0.001]
Observations	27,072	876,865	37,044	866,893
Depvar Mean	0.333	0.0341	0.326	0.0377
LL, Non-MF Mean	0.4	0.0428	0.385	0.049
MF + MF $\times$ $LH = 0$ p-val	0.039	0.002	0.025	0.001
MF + MF $\times$ $HH = 0$ p-val	0.066	0.006	0.03	0.002
MF + $LH\times$ MF = MF + $HH\times$ MF p-val	0.955	0.325	0.838	0.326

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Columns 1-2 restrict to financial links, while columns 3-4 restrict to non-financial links. Columns 1 and 3 consider links that existed in Wave 1, while columns 2 and 4 consider pairs of nodes that were not linked in Wave 1.

Table J.4. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance	−0.105	−0.109	−0.077	−0.077
	(0.032)	(0.031)	(0.022)	(0.020)
	[0.002]	[0.0004]	[0.0005]	[0.0002]
Microfinance × LLH	0.063	0.047	0.025	0.013
	(0.024)	(0.022)	(0.019)	(0.016)
	[0.011]	[0.036]	[0.197]	[0.432]
Microfinance × LHH	0.077	0.057	0.034	0.016
	(0.032)	(0.027)	(0.030)	(0.024)
	[0.015]	[0.034]	[0.248]	[0.505]
Microfinance × HHH	0.090	0.064	0.046	0.022
	(0.038)	(0.030)	(0.027)	(0.023)
	[0.019]	[0.036]	[0.093]	[0.335]
LLH	−0.083	−0.064	−0.045	−0.034
	(0.020)	(0.020)	(0.012)	(0.013)
	[0.00004]	[0.002]	[0.0002]	[0.008]
LHH	−0.098	−0.072	−0.060	−0.045
	(0.026)	(0.026)	(0.017)	(0.019)
	[0.0002]	[0.005]	[0.001]	[0.017]
HHH	−0.075	−0.059	−0.020	−0.017
	(0.032)	(0.032)	(0.017)	(0.024)
	[0.020]	[0.067]	[0.229]	[0.470]
Observations	53,233	53,233	53,233	53,233
Depvar Mean	0.197	0.197	0.808	0.808
LLL, Non-MF Mean	0.297	0.297	0.876	0.876
MF + MF × $HHH = 0$ p-val	0.596	0.051	0.101	0.003
MF + MF × $LLH = 0$ p-val	0.046	0.004	0.004	0.001
MF + MF × $LHH = 0$ p-val	0.219	0.017	0.063	0.003
MF + MF × HHH = MF + MF × LLH p-val	0.323	0.432	0.274	0.591
MF + MF × HHH = MF + MF × LHH p-val	0.538	0.658	0.419	0.651
MF + MF × LLH = MF + MF × LHH p-val	0.365	0.495	0.574	0.826

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

Table J.5. Borrowing patterns

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance × Post	657.376	−528.719	−837.834	−395.597	491.097
	(148.561)	(357.216)	(424.268)	(778.745)	(681.164)
	[0.00001]	[0.139]	[0.049]	[0.612]	[0.471]
Microfinance × Post × H	739.096	254.575	−31.136	302.433	−833.325
	(241.762)	(270.868)	(305.040)	(901.182)	(1,080.399)
	[0.003]	[0.348]	[0.919]	[0.738]	[0.441]
Microfinance × H	109.430	23.072	−202.299	319.122	554.483
	(56.906)	(41.676)	(179.781)	(479.307)	(720.250)
	[0.055]	[0.580]	[0.261]	[0.506]	[0.442]
Post × H	310.332	−633.258	101.327	−330.929	−342.200
	(142.121)	(220.026)	(233.979)	(710.181)	(631.432)
	[0.029]	[0.005]	[0.665]	[0.642]	[0.588]
Observations	28,062	27,194	28,062	28,062	28,062
Depvar Mean	596.976	860.228	1863.324	2667.56	1656.881
L, Non-MF Mean	178.436	1227.156	1914.277	2831.144	1845.938
MF × Post ×H + MF × Post =0 p-val	0.000	0.326	0.053	0.92	0.679

Notes: Classification of H type is based on logistic regression. This table presents the effect of microfinance access on the loan amounts borrowed from various sources. Outcomes are winsorized to the 1% level. All of the columns control for surveyed in wave 1 fixed effects. Here all specifications include demographic household and village controls (the same ones used in random forest classification of H vs L) subject to double-post LASSO. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. MFI: Microfinance Institution; SHG: Self-Help Group

Table J.6. Confusion Matrices for H and L classification, Karnataka

		Predicted L	Predicted H	Total
Observed	L	866	1501	2367
	H	84	428	512
	Total	950	1929	$N = 2879$

Notes: This table presents the confusion matrix for the validation sample for Karnataka. The following metrics on this confusion matrix capture classification quality: DOR = 2.94, F1 = 0.351, MCC = 0.164.

J.2. **Hyderabad Exhibits.**

Table J.7. Characteristics of H versus L, Hyderabad

Panel A: Hyderabad - Demographics and Amenities variables

	(1)	(2)	(3)	(4)	(5)
	GMOBC	Latrine	Num. Rooms	Thatched Roof	RCC Roof
H	0.024	0.072	0.193	−0.003	−0.027
	(0.025)	(0.019)	(0.081)	(0.007)	(0.018)
	[0.351]	[0.000]	[0.019]	[0.690]	[0.141]
Depvar Mean	0.429	0.578	2.314	0.025	0.882
Observations	4,520	4,483	4,516	4,516	4,508

Panel B: Hyderabad - Network Variables

	(1)	(2)	(3)	(4)
	Expected Degree	Expected Links to L	Expected Links to H	Expected Centrality
H	0.373	0.100	0.272	0.010
	(0.123)	(0.123)	(0.090)	(0.003)
	[0.003]	[0.418]	[0.003]	[0.001]
Depvar Mean	5.806	4.144	1.664	0.074
Observations	4,523	4,523	4,523	4,523

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. GMOBC = A dummy for whether the household consists of general caste, otherwise back-wards caste, so the omitted categories are scheduled caste and scheduled tribes. General andOBC are considered upper caste. RCC is Reinforced Cement Concrete.

Table J.8. Link Evolution for Financial and Non Financial Links, Hyderabad

	(1) Financial Links	(2) Financial Links	(3) Non Financial Links	(4) Non Financial Links
Microfinance	−0.364	−0.311	−0.283	−0.188
	(0.112)	(0.116)	(0.089)	(0.081)
	[0.002]	[0.009]	[0.003]	[0.024]
Microfinance × H	0.401	0.448	0.637	0.666
	(0.157)	(0.153)	(0.180)	(0.180)
	[0.013]	[0.005]	[0.001]	[0.000]
H	0.167	0.021	−0.007	−0.100
	(0.120)	(0.123)	(0.132)	(0.131)
	[0.167]	[0.862]	[0.961]	[0.449]
Observations	4,429	4,429	4,429	4,429
Double-Post LASSO	No	Yes	No	Yes
Depvar Mean	4.24	4.24	2.87	2.87
MF + MF × H = 0 p-val	0.824	0.42	0.051	0.005

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. All columns include a full set of controls. Centrality controls are a vector of flexible controls (a polynomial) for centrality of both nodes. Household characteristics are caste and a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material. Household predictor variables consist of all variables that are used in the random forest classification. In every case we include interactions of all of these network, demographic, and classification variables with microfinance.

Table J.9. Link Evolution, Hyderabad

	(1) Prob. Linked	(2) Prob. Linked
Microfinance	−0.005	−0.006
	(0.002)	(0.003)
	[0.056]	[0.012]
Microfinance x HH	−0.008	−0.008
	(0.004)	(0.004)
	[0.074]	[0.037]
Microfinance x LH	−0.001	−0.001
	(0.001)	(0.00)
	[0.354	[0.355]
HH	0.013	0.013
	(0.004)	(0.003)
	[0.002]	[0.0003]
LH	0.004	0.004
	(0.001)	(0.001)
	[0.004]	[0.002]
Observations	141,996	141,996
Controls	No	Yes
Depvar Mean	0.0255	0.0255
LL, Non MF Mean	0.0257	0.0257
MF + MF x $HH = 0$ p-val	0.020	0.002
MF + MF x $LH = 0$ p-val	0.049	0.010
MF + MF x HH = MF + MF x LH p-val	0.037	0.012

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. The controls are selected by double post lasso among all the variables that are used for its random forest classification, and includes several household and village level characteristics.

Table J.10. Triples Evolution, Hyderabad

	(1)	(2)
All variables x 1000	Full Triangle Linked	Full Triangle Linked
Microfinance	-0.02	-0.04
	(0.01)	(0.02)
	[0.164]	[0.109]
Microfinance $\times$ LLH	-0.01	-0.01
	(0.01)	(0.01)
	[0.266]	[0.298]
Microfinance $\times$ LHH	-0.04	-0.03
	(0.02)	(0.02)
	[0.087]	[0.044]
Microfinance $\times$ HHH	-0.1	-0.1
	(0.1)	(0.05)
	[0.040]	[0.015]
LLH	0.01	0.02
	(0.01)	(0.01)
	[0.026]	[0.017]
LHH	0.05	0.1
	(0.02)	(0.02)
	[0.020]	[0.007]
HHH	0.2	0.2
	(0.1)	(0.1)
	[0.017]	[0.006]
Observations	3,341,006	3,341,006
Controls	No	Yes
Depvar Mean	3.53e-02	3.53e-02
LLL, Non-MF Mean	2.99e-02	2.99e-02
MF + MF x HHH = 0 p-val	0.03	0.006
MF + MF x LLH = 0 p-val	0.139	0.103
MF + MF x LHH = 0 p-val	0.058	0.024
MF + MF x HHH = MF + MF x LLH p-val	0.038	0.015
MF + MF x HHH = MF + MF x LHH p-val	0.034	0.015
MF + MF x LLH = MF + MF x LHH p-val	0.072	0.032

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. The controls are all the variables that are used for its random forest classification, and includes several household and village level characteristics.

Table J.11. Borrowing patterns, Hyderabad

	(1) MFI	(2) Friends	(3) SHG	(4) Moneylender	(5) Family
Microfinance	1,161.140	−43.085	−1,938.432	−1,494.645	41.946
	(282.913)	(821.081)	(905.413)	(1,548.317)	(685.827)
	[0.0001]	[0.959]	[0.035]	[0.337]	[0.952]
Microfinance × H	1,947.747	93.715	−95.705	299.438	847.565
	(488.351)	(1,348.892)	(1,553.800)	(2,385.734)	(1,126.558)
	[0.0002]	[0.945]	[0.952]	[0.901]	[0.454]
H	725.643	−5.442	2,377.161	155.550	321.567
	(265.028)	(1,127.431)	(1,220.323)	(1,734.179)	(813.402)
	[0.008]	[0.997]	[0.055]	[0.929]	[0.694]
Observations	6,811	6,863	6,863	6,863	6,863
Depvar Mean	3107.86	7895.05	6935.66	18805.06	2620.97
L, Non MF Mean	2079.1	7895.74	7020.19	19372.79	2634.7
MF + MF × H = 0 p-val	0	0.97	0.221	0.647	0.478

Notes: Classification of H type is based on logistic regression. These tables present the effect of microfinance access on the loan amounts borrowed from various sources. Outcomes are winsorized to the 2.5% level. Here all specifications include demographic household and village controls (the same ones used in random forest classification of H vs L) subject to double-post LASSO. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Table J.12. Confusion Matrices for H and L classification, Hyderabad

		Predicted L	Predicted H	Total
Observed	L	651	184	835
	H	153	81	234
	Total	804	265	$N = 1069$

Notes: Classification of H type is based on logistic regression. This table presents the confusion matrix for the validation sample for Hyderabad. The following metrics on this confusion matrix capture classification quality: DOR = 1.87, F1 = 0.325, MCC = 0.120.

TABLE J.13. Risk sharing, Hyderabad

	(1) Expenditures: Non-Food	(2) Expenditures: Total
Microfinance × Income	0.079	0.071
	(0.030)	(0.037)
	[0.009]	[0.056]
Microfinance × Income $\times H$	−0.076	−0.096
	(0.045)	(0.059)
	[0.097]	[0.107]
Household Income per capita	0.054	0.113
	(0.020)	(0.026)
	[0.011]	[0.000]
Household Income per capita $\times H$	0.025	0.041
	(0.028)	(0.043)
	[0.389]	[0.347]
Observations	10,502	10,593
Depvar Mean	1193	2040
L, Non-MF Depvar Mean	1184	2055
Income Mean	1440	1437
L, Non-MF Income Mean	1437	1434
Test: MF x Income + MF x Income x H = 0	0.931	0.618

Notes: Classification of H type is based on logistic regression. Income is total household, monthly per capita earnings from employment or business activities, excluding private and government transfers. Dependent variable is monthly per capita household expenditure. In col. 1, expenditure excludes food and in col. 2, we present total expenditure. Data is from the first (2007-08) and third (2012) waves of the Hyderabad survey. Regression includes controls for household fixed effects and wave-by-neighborhood-by-type fixed effects. Additional controls are selected by double post lasso from the set of variables used in the prediction exercise, interacted with type. Standard errors (clustered at the neighborhood level) are reported in parentheses. p-values are reported in brackets.

Appendix K. Karnataka Robustness

Table K.1. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF	(5) Linked Post-MF	(6) Linked Post-MF
Microfinance	−0.058	−0.038	−0.033	−0.023	−0.006	−0.005
	(0.018)	(0.018)	(0.017)	(0.008)	(0.006)	(0.006)
	[0.002]	[0.031]	[0.058]	[0.006]	[0.261]	[0.431]
Microfinance × LH	0.009	0.007	0.002	0.007	0.006	0.005
	(0.015)	(0.015)	(0.017)	(0.004)	(0.003)	(0.003)
	[0.573]	[0.627]	[0.931]	[0.120]	[0.070]	[0.141]
Microfinance × HH	0.039	0.038	0.018	0.009	0.007	0.010
	(0.022)	(0.022)	(0.020)	(0.007)	(0.005)	(0.005)
	[0.086]	[0.077]	[0.358]	[0.206]	[0.183]	[0.082]
LH	−0.025	−0.030	−0.051	−0.002	−0.006	−0.013
	(0.012)	(0.012)	(0.025)	(0.004)	(0.003)	(0.006)
	[0.036]	[0.016]	[0.046]	[0.566]	[0.027]	[0.039]
HH	0.008	−0.001	−0.065	0.021	0.013	0.015
	(0.017)	(0.016)	(0.046)	(0.006)	(0.004)	(0.010)
	[0.622]	[0.934]	[0.155]	[0.001]	[0.002]	[0.130]
Observations	57,376	57,376	57,376	846,561	846,561	846,561
Linked Pre-MF	Yes	Yes	Yes	No	No	No
Village size control		✓	✓		✓	✓
Village size interacted with LH,HH control			✓			✓
Depvar Mean	0.441	0.441	0.441	0.0636	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.482	0.0753	0.0753	0.0753
MF + MF × $LH = 0$ p-val	0.014	0.154	0.176	0.015	0.901	0.996
MF + MF × $HH = 0$ p-val	0.361	0.996	0.51	0.101	0.927	0.475
MF + LH × MF = MF + HH × MF p-val	0.137	0.118	0.415	0.641	0.788	0.31

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Village size interactions are village size interacted with with LLH, LHH and HHH respectively.

TABLE K.2. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance × LH	0.009	−0.004	0.007	0.008
	(0.015)	(0.015)	(0.004)	(0.005)
	[0.573]	[0.796]	[0.120]	[0.077]
Microfinance × HH	0.039	0.015	0.009	0.014
	(0.022)	(0.024)	(0.007)	(0.007)
	[0.086]	[0.512]	[0.206]	[0.039]
LH	−0.025	−0.002	−0.002	−0.008
	(0.012)	(0.012)	(0.004)	(0.004)
	[0.036]	[0.864]	[0.566]	[0.068]
HH	0.008	0.046	0.021	0.006
	(0.017)	(0.020)	(0.006)	(0.007)
	[0.622]	[0.023]	[0.001]	[0.358]
Observations	57,376	57,376	846,561	846,561
Linked Pre-MF	Yes	Yes	No	No
Controls		✓		✓
Depvar Mean	0.441	0.441	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.0753	0.0753
MF + MF $\times LH = 0$ p-val	0.014	0.18	0.015	0.843
MF + MF $\times HH = 0$ p-val	0.361	0.244	0.101	0.652
MF $+LH\times$ MF = MF + $HH\times$ MF p-val	0.137	0.344	0.641	0.195

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls (interacted with MF) are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

Table K.3. Triple Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Full triangle linked Post-MF	(4) Any link in triangle survived Post-MF	(5) Any link in triangle survived Post-MF	(6) Any link in triangle survived Post-MF
Microfinance	−0.078	−0.063	−0.054	−0.085	−0.069	−0.056
	(0.029)	(0.030)	(0.030)	(0.023)	(0.023)	(0.024)
	[0.008]	[0.039]	[0.071]	[0.0002]	[0.003]	[0.020]
Microfinance × LLH	0.026	0.024	0.023	0.043	0.042	0.030
	(0.021)	(0.022)	(0.025)	(0.018)	(0.018)	(0.021)
	[0.228]	[0.266]	[0.362]	[0.015]	[0.020]	[0.154]
Microfinance × LHH	0.054	0.052	0.029	0.057	0.055	0.031
	(0.030)	(0.030)	(0.029)	(0.025)	(0.024)	(0.025)
	[0.072]	[0.083]	[0.315]	[0.022]	[0.024]	[0.223]
Microfinance × HHH	0.093	0.093	0.074	0.087	0.087	0.053
	(0.042)	(0.042)	(0.041)	(0.031)	(0.031)	(0.033)
	[0.028]	[0.026]	[0.075]	[0.006]	[0.006]	[0.108]
LLH	−0.024	−0.028	−0.034	−0.037	−0.040	−0.077
	(0.018)	(0.019)	(0.032)	(0.014)	(0.015)	(0.023)
	[0.180]	[0.146]	[0.294]	[0.009]	[0.006]	[0.001]
LHH	−0.037	−0.043	−0.119	−0.032	−0.038	−0.114
	(0.025)	(0.025)	(0.047)	(0.017)	(0.017)	(0.033)
	[0.133]	[0.084]	[0.012]	[0.053]	[0.022]	[0.001]
HHH	−0.025	−0.034	−0.089	−0.012	−0.021	−0.120
	(0.033)	(0.032)	(0.096)	(0.022)	(0.022)	(0.055)
	[0.454]	[0.289]	[0.353]	[0.593]	[0.323]	[0.030]
Observations	53,233	53,233	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes	Yes	Yes
Village size control		✓	✓		✓	✓
Village size interacted with LLH,LHH,HHH control			✓			✓
Depvar Mean	0.197	0.197	0.197	0.808	0.808	0.808
LLL, Non-MF Mean	0.252	0.252	0.252	0.864	0.864	0.864
MF + MF × HHH = 0 p-val	0.698	0.428	0.632	0.935	0.444	0.908
MF + MF × LLH = 0 p-val	0.023	0.119	0.211	0.022	0.155	0.186
MF + MF × LHH = 0 p-val	0.262	0.665	0.303	0.141	0.498	0.264
MF + MF × HHH = MF + MF × LLH p-val	0.076	0.066	0.213	0.093	0.083	0.406
MF + MF × HHH = MF + MF × LHH p-val	0.212	0.184	0.182	0.075	0.06	0.266
MF + MF × LLH = MF + MF × LHH p-val	0.122	0.127	0.733	0.409	0.433	0.957

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Village size interactions are village size interacted with with LLH, LHH and HHH respectively.

Table K.4. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance × LLH	0.026	−0.002	0.043	0.023
	(0.021)	(0.022)	(0.018)	(0.016)
	[0.228]	[0.946]	[0.015]	[0.149]
Microfinance × LHH	0.054	−0.004	0.057	0.021
	(0.030)	(0.029)	(0.025)	(0.019)
	[0.072]	[0.896]	[0.022]	[0.282]
Microfinance × HHH	0.093	0.002	0.087	0.037
	(0.042)	(0.040)	(0.031)	(0.028)
	[0.028]	[0.953]	[0.006]	[0.194]
LLH	−0.024	0.007	−0.037	−0.018
	(0.018)	(0.020)	(0.014)	(0.014)
	[0.180]	[0.718]	[0.009]	[0.203]
LHH	−0.037	0.030	−0.032	0.006
	(0.025)	(0.025)	(0.017)	(0.014)
	[0.133]	[0.242]	[0.053]	[0.696]
HHH	−0.025	0.076	−0.012	0.039
	(0.033)	(0.032)	(0.022)	(0.021)
	[0.454]	[0.020]	[0.593]	[0.068]
Observations	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes
Controls		✓		✓
Depvar Mean	0.197	0.197	0.808	0.808
LLL, Non-MF Mean	0.252	0.252	0.864	0.864
MF + MF x HHH = 0 p-val	0.698	0.303	0.935	0.762
MF + MF x LLH= 0 p-val	0.023	0.273	0.022	0.686
MF + MF x LHH= 0 p-val	0.262	0.28	0.141	0.666
MF + MF x HHH = MF + MF x LLH p-val	0.076	0.909	0.093	0.604
MF + MF x HHH = MF + MF x LHH p-val	0.212	0.83	0.075	0.369
MF + MF x LLH= MF + MF x LHH p-val	0.122	0.894	0.409	0.868

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Controls (interacted with MF) are selected by double post lasso among centrality controls (vector of flexible controls for centrality of both nodes), household characteristics (caste, a number of wealth proxies including number of rooms, number of beds, electrification, latrine presence, and roofing material) and all variables that are used in the random forest classification.

Appendix L. Hyderabad Consumption Smoothing Robustness

Table L.1. Microfinance Treatment Effects on Income, Hyderabad

	(1) Income
Microfinance	-23.429
	(63.771)
	[0.713]
Microfinance $\times H$	4.136
	(83.417)
	[0.960]
Observations	10457.000
Depvar Mean	1436.747
L, Non MF Mean	1434.223
Test: MF + MF $\times H$	0.786

Notes: Income is total household, monthly per capita earnings from employment or business activities. Regression includes controls for strata fixed effects. Additional controls are selected by double post lasso from the set of variables used in the prediction exercise, interacted with type. Standard errors (clustered at the neighborhood level) are reported in parentheses. p-values are reported in brackets.

Appendix M. Model Simulation

Here we provide simulation evidence that, in our model, LL links can drop more than HL or HH links in response to an exogenous change in relationship value. We consider the model specialized to the case of two types, as in Section 4.6.

In our simulations, we maintain the following parameters:

- $n = 250$; population
- $\lambda_H = 0.2$; the share of high types, which approximates the empirical frequency
- $\lambda_L = (1 - \lambda_H)$; the share of low types
- $F(\cdot)$ is the uniform distribution on [0,1]
- values for links
  - $v_{HH} = 0.46$
  - $v_{LL} = 0.27$
  - In our simulations we vary v_{HL} and v_{LH} over the range

    $$(v_{HL}, v_{LH}) \in [0.05, 0.5]^2$$

    to trace out relative changes in linking effort across types over the parameter space.
- base costs and benefits of socializing
  - $c = 4$; homogenous cost
  - $u = 0.3$; homogenous benefit

We show below the resulting stylized network is sparse and homophilic, much like the data. Nevertheless, in the homophilic network, we will find a non-trivial part of the parameter space where declines of effort are greater for Ls than Hs and where links decline more between two L types than any other configuration of types.

As shown in the paper, unique equilibrium efforts can be calculated as

$$e = (I - E)^{-1} u$$

with E being the $|\Theta| \times |\Theta|$ matrix with θ, θ' entries

$$\frac{1}{c_\theta} \mathrm{E}^+[v_{\theta\theta'}] n_{\theta\theta'} \left(1 - F(-v_{\theta\theta'})\right) \left(1 - F(-v_{\theta'\theta})\right).$$

Further, as shown in Section 4.6, we can directly calculate the expected degrees of nodes of each type, d_θ, as well as the link counts from θ to θ', denoted $d_{\theta\theta'}$.

We are interested in how e changes with the introduction of microcredit. Microfinance is assumed to affect valuation only through one parameter; it reduces v_{HL}:

$$v_{HL}^{\text{mf}} = 0.75 v_{HL}^{\text{no mf}}$$
$$v_{LH}^{\text{mf}} = v_{LH}^{\text{no mf}}$$
$$v_{LL}^{\text{mf}} = v_{LL}^{\text{no mf}}$$

$$v_{HH}^{\text{mf}} = v_{HHH}^{\text{no mf}}.$$

By varying the valuation of cross-type links, we can study how the equilibrium efforts change across the $(v_{HL}^{\text{no mf}}, v_{LH})$-plane. In what follows, v_{HL} denotes $v_{HL}^{\text{no mf}}$.

Specifically, we look the quantity

$$\delta(v_{HL}, v_{LH}) := \frac{e_H^{\text{mf}} - e_H^{\text{no mf}}}{e_L^{\text{mf}} - e_L^{\text{no mf}}}$$

which is a positive ratio, since both efforts decline due to microcredit as per Proposition 1.

Figure M.1 plots $\log \delta$ as a function of (v_{HL}, v_{LH}). Because δ is a ratio of changes in fractions, the logarithmic transformation makes visualization easier by dampening extremes for visualization. So, notice that if $\log \delta < 0$, then the decline in effort for H types is less than that of L types which is the pathology made possible under this model.

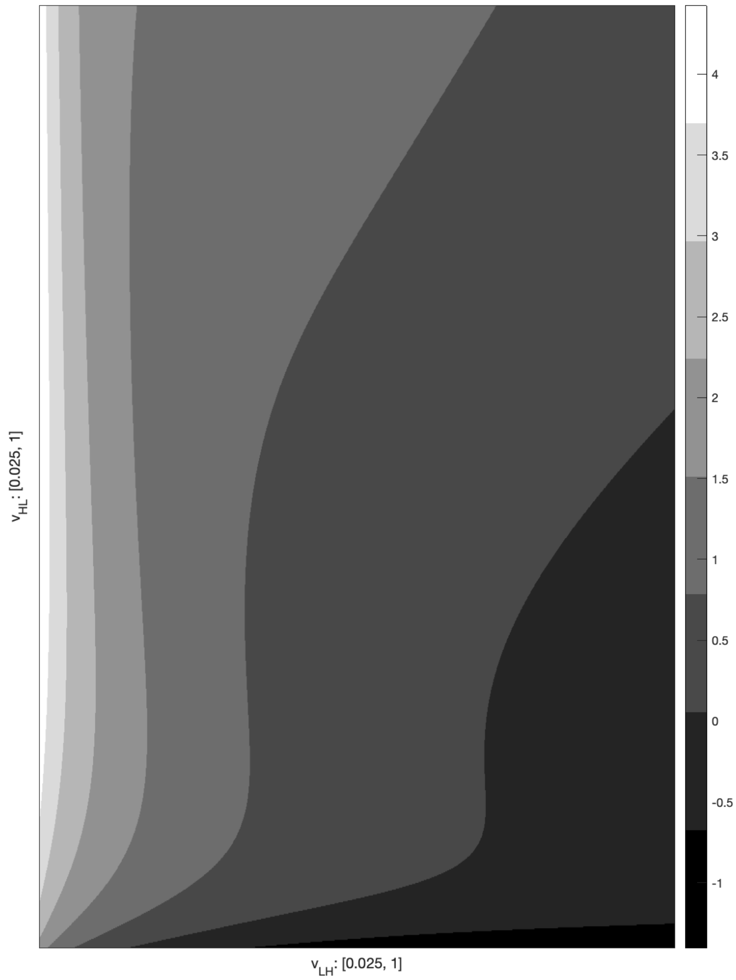


Figure M.1. Plot of $\log \delta$ as a function of (v_{HL}, v_{LH})

From the figure it is clear that effort declines more for Ls when $v_{LH} >> v_{HL}$ (bottom, right area of the parameter space). Moreover, for larger values of v_{LH}, we see larger declines in e_L than e_H (or $\log \delta < 0$), even for larger values of v_{HL}. Further, the intuition is that when Hs place lower value on Ls, but Ls still value Hs quite a bit, then microcredit greatly reduces mutual consent. As a consequence, much of the motivation on behalf of the Ls themselves globally reduces and it can result in greater declines in effort for Ls.

In Table M.1, we present the basic link patterns for an example in the parameter space, with $v_{HL} = 0.01$ and $v_{LH} = 0.8$. With the chosen parameters, the network is sparse (an average degree of roughly 9) but also exhibits considerable homophily. The vast majority of links are within type, rather than across type. This also points to the fact that just because Ls value Hs more does not mean that the network need to be heterophilic. It is useful to note that links require mutual consent, so even if Ls highly value Hs, the fact that Hs do not value Ls much can generate homophily. This is amplified by the fact that $\lambda_H = 0.2$, so, as there are few Hs to begin with, the resulting network would still be homophilic.

In this example we see a sparse network, with strong homophily, with uniform declines in link patterns of all types and efforts in linking when microfinance is introduced. Further, the declines for d_{LL} are the largest and also e_L declines more than e_H.

TABLE M.1. Link and Effort Patterns for an Example

	No MF	MF	Difference
d_H	9.614	9.287	-0.327
d_L	8.689	8.226	-0.463
e_H	0.898	0.896	-0.002
e_L	0.760	0.742	-0.018
d_{HH}	8.523	8.490	-0.033
d_{HL}	1.091	0.798	-0.293
d_{LH}	0.273	0.199	-0.073
d_{LL}	8.416	8.027	-0.389

Appendix N. Karnataka Links to New Households

Table N.1. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF
Microfinance	−0.030	−0.024
	(0.009)	(0.010)
	[0.002]	[0.021]
HL x MF	0.009	0.007
	(0.005)	(0.007)
	[0.053]	[0.268]
HH x MF	0.013	−0.006
	(0.010)	(0.010)
	[0.197]	[0.529]
HL	−0.005	−0.004
	(0.004)	(0.006)
	[0.190]	[0.446]
HH	0.033	0.042
	(0.008)	(0.009)
	[0.000]	[0.000]
Observations	577,872	326,065
Link to a new in-migrated house	Yes	No
Depvar Mean	0.0897	0.0838
LL, Non-MF Mean	0.104	0.0961
MF + MF x $HL = 0$ p-val	0.014	0.027
MF + MF x $HH = 0$ p-val	0.147	0.002
MF + HL x MF = MF + HH x MF p-val	0.66	0.078

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

Appendix O. Effects of microfinance: Example

Consider the special case where $\alpha_L = \alpha_H$, $\beta_L = \beta_H$ and $\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H = 0$. In this case $\Delta_{HH} = 0$.

Now suppose first that $\Delta\beta_H > 0$ and therefore $\Delta\alpha_H < 0$. In this case

$$0 < \mathit{\Delta}_{HL} = \alpha_H \Delta\beta_H b + \beta_H \Delta\alpha_H r \Leftrightarrow r\frac{\beta_H|\Delta\alpha_H|}{\alpha_H \Delta\beta_H} = r < b$$

and

$$0 < \mathit{\Delta}_{LH} = \alpha_H \Delta\beta_H r + \beta_H \Delta\alpha_H b \Leftrightarrow r\frac{\alpha_H \Delta\beta_H}{\beta_H|\Delta\alpha_H|} = r > b.$$

In the case where $\Delta\beta_H < 0$ and therefore $\Delta\alpha_H > 0$, these inequalities get reversed and we get

$$0 < \Delta_{HL} \Leftrightarrow r > b$$

and

$$0 < \Delta_{LH} \Leftrightarrow r < b.$$

In other words, in this special case, Δ_{HL} and Δ_{LH} move in opposite directions and which one goes up depends on which of r and b is bigger and whether or not $\Delta\beta_H > 0$.

Since $b > r$, in this special example we would expect Δ_{HL} to be positive and Δ_{LH} to be negative as long as $\Delta\beta_H > 0$ and the reverse otherwise. In other words, it is entirely possible for v_{HL} to go up, v_{LH} to go down and v_{HH} to be unchanged but it requires α_H to go down and β_H to go up.