Review of Economic Studies (2024) **91**, 1331–1372 https://doi.org/10.1093/restud/rdad065

Advance access publication 27 June 2023

# Changes in Social Network Structure in Response to Exposure to Formal Credit Markets

Abhijit Banerjee
*Department of Economics, MIT, Cambridge and NBER, Cambridge and JPAL, Cambridge USA*
Emily Breza
*Department of Economics, Harvard University, Cambridge and NBER, Cambridge and JPAL, Cambridge, USA*
Arun G. Chandrasekhar
*Department of Economics, Stanford University, Stanford and NBER, Cambridge and JPAL, Cambridge, USA*
Esther Duflo
*Department of Economics, MIT, Cambridge USA and Collège de France, Paris France*
Matthew O. Jackson
*Department of Economics, Stanford University, Stanford and Santa Fe Institute, Santa Fe USA*
and
Cynthia Kinnan
*Department of Economics, Tufts University, Medford and NBER, Cambridge and JPAL, Cambridge USA*

*First version received September* 2018*; Editorial decision June* 2022*; Accepted June* 2023 *(Eds.)*

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 randomised controlled trial (RCT) in Hyderabad, where a microfinance institution randomly selected 52 of 104 neighbourhoods to enter first. Four years after all neighbourhoods were treated, households in early-entry neighbourhoods had credit access longer and had larger loans. We again find fewer social relationships between households in these neighbourhoods, 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 analyse 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.

*The editor in charge of this paper was Adam Szeidl.*

*Key words*: Social networks, Network change, Network formation, Network evolution, Microfinance, Market exposure, Favor exchange, Social capital

*JEL codes*: D85, D13, L14, O12, Z13

## 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 analyse 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 MF 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 analyse the introduction of MF in rural Karnataka, India using two waves of detailed network panel data that we collected (Banerjee *et al*., 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 MF, but an MFI, 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 households) to estimate the impact of MF on village network structures using a difference-in-difference strategy.

Second, we replicate and extend the Karnataka findings, leveraging an RCT conducted in 104 neighbourhoods in Hyderabad, India (Banerjee *et al*., 2015a, 2019a). In the RCT, entry by an MFI (Spandana) was randomized to half of the study neighbourhoods. 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

1. See, *e.g.* Udry (1994), Fafchamps and Lund (2003), Karlan *et al*. (2009), Beaman and Magruder (2012), Ambrus *et al*. (2014), Blumenstock *et al*. (2016), Munshi and Rosenzweig (2016), Blumenstock and Tan (2016) and Breza (2016).

six years after initial entry of microcredit. At this point, they had little or no access to microcredit, though households in the early entry neighbourhoods 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 preperiod network structures. 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 neighbourhoods had exogenously more cumulative access to MF than control neighbourhoods. 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 these 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 are 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.* (2019, 2020) 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 *et al.* (2019, 2020) 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 neighbourhood network that we need for our analysis. Breza *et al.* (2019, 2020) 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 MF on network connections for those involved can potentially go in either direction. As a source of formal credit to poor, underbanked households, MF may reduce dependence on social networks for informal credit and insurance, and this effect may be heterogeneous across households (Islam *et al.*, 2015). Moreover, the required weekly repayment structure of microloans may reduce borrowers' liquidity and limit their capacity to lend small sums to their friends (Field *et al.*, 2012). On the other hand, if households re-lend a part of their MF loans, MF could crowd in informal financial relationships.[2]

In both of our datasets, we find that the introduction of MF 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 MF loans and those who are unlikely to do so. All of the channels described above suggest that MF might affect borrowers' willingness to maintain friendships, including with those who do not take up MF. 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 MF. If anything, one would have expected links between

2. Kinnan and Townsend (2012), Field *et al.* (2012), Feigenberg *et al.* (2013), and Vera-Cossio (2019) find evidence consistent with the households re-lending bank and credit cooperative loans.

these households to be *strengthened* in MF villages, since they might be losing access to the households that get MF 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 versus less likely to take up MF in MF villages/neighbourhoods to those in a non-MF village who would have been comparably likely to take up MF had it been available in their village/neighbourhood. 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 MF 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*, and *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 Online Appendix E, which largely predict that we should see exposure effects that parallel the direct impact of treatment.

Specifically, in the Karnataka panel, when we examine the probability that two *L*s 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 deliver consistent results. Treated MF neighbourhoods have 0.7pp (22%) fewer *LL* links than control neighbourhoods ($p = 0.004$), 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, *LL*s and *LLL*s 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 Online Appendix C, that number climbs to 50%.

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.

The evidence suggests that in Karnataka, *LL*s and *LLL*s 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 *H*s. However, the fact that we see *greater* drops in Karnataka means that whatever dynamics underpin the behavioural response by individuals must allow for these "non-monotone" effects.

Moreover, the fact that *LL*s and *LLL*s 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 MF is likely to be on financial links, the same patterns also emerge when we analyse information (*i.e.* advice-giving and -receiving) links. In Hyderabad, the evidence for this phenomenon is more suggestive among *L*s, but a similar multiplexing spillover is seen for the *H*s. 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 E: 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 *L*s 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 *L*s can lose at least as much as *H*s. 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 MF 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 *H*s becoming less willing to lend can have a larger negative impact on *L*s than on *H*s, which leads to less socializing by *L*s. As *L*s 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

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*. (2010), König *et al*. (2014), Currarini *et al*. (2009), Currarini *et al*. (2010), Cabrales *et al*. (2011), and Canen *et al*. (2017).

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 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 *L*s, 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 MF, 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 MF exposure on consumption smoothing for *H* versus *L* households. This is possible because we have detailed household-level panel information on both income and consumption. In areas exposed to MF, households with high propensity to use MF (*H*s) see little change in their consumption smoothing compared to those in areas not exposed to MF. However, households with low propensity to use microfinance (*L*s) 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 *et al*. (2013) and Comola and Prina (2017) explore whether and in what ways financial interventions affect participating households' networks.[5]

In recent work, Heß *et al*. (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

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 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.

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. Proofs, robustness exercises, and extensions can be found in the Online Appendix and the Auxiliary Appendix (available on the authors' webpages).

## 2. SETTING, DATA, AND SAMPLE STATISTICS

### 2.1. *Setting*

**2.1.1. Karnataka (India).** In 2006, the MF 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 five districts of the state of Karnataka in India. Prior to BSS's entry, these villages had minimal exposure to MF.

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 MFI—randomly chose 52 of 104 neighbourhoods in Hyderabad (at the time the capital of Andhra Pradesh, a State neighbouring Karnataka, in South India) to enter. After two years, the remaining 52 neighbourhoods received

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

access in mid-2008. The short- and medium-run impacts of randomized access to MF 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 neighbourhoods had been exposed to microcredit for 6 years (4 years of active lending) and the control neighbourhoods had been exposed for 3.5 years (1.5 years of active lending). Network data were collected during this third round.

The early treatment neighbourhoods had greater MF access overall. As MF 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 neighbourhoods still had 14% more contemporaneous MF borrowing and 43% more cumulative MF borrowing over the preceding three years (Banerjee *et al*., 2019a). However, since nobody had access to MF 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*

**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

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.

8. The Wave 1 data are described in detail in Banerjee *et al*. (2013), and both waves are publicly available at http://economics.mit.edu/faculty/eduflo/social.

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) $= 1-(1-0.46)^2 = 0.708$.

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, MFI, 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 MF exposure, and 4.8% Wave 2 households in-migrated (which we cannot use in the panel) or split off from existing households (as children reach adulthood), again not differential by MF 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, 2019a). The first round of data collection was conducted in late 2007–early 2008, 15–18 months after MF was made available in the treatment group. Following this first wave, the control group also received access to MF in May 2008. A second round of data collection was conducted in mid-2010 to examine longer-term impacts of access to MF; 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 MF 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 neighbourhoods in Hyderabad, we chose to collect Aggregated Relational Data (ARD), which is used as described below. As 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 neighbourhood in the Hyderabad sample were surveyed and asked a set of network questions. First, respondents were asked how many links they had within the neighbourhood (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 neighbourhood 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?" Online Appendix F.1 details both types of survey questions. Third, we asked each sampled household whether they themselves possessed each of the ARD traits.

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).

We follow Breza *et al*. (2020) 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 $A$s than $B$s. Further, if individual $i$ is of type $C$, then this means that $C$s are more likely to link to $A$s than to $B$s. In that sense, ARD allows the researcher, under appropriate parametric assumptions, to model the underlying network structure despite not having collected link-level data, which can then be used in econometric estimation. Breza *et al*. (2020) develop this method, and we also provide further details in Online Appendix F.2.

A rough sketch is as follows. We assume a standard parametric model of network formation (Hoff *et al*., 2002). There is some (unobserved) latent space where households reside and every household has some location $z_i$ as well as a fixed effect $\nu_i$, 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 ($\zeta_k$) 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*. (2020, 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*. (2020) 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 have information about one single type of link encompassing all dimensions of interaction, both financial and non-financial.

12. For example, Breza *et al*. (2020) 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.

TABLE 1
*Sample statistics*

| | Obs | Control mean | Control SD | Treatment—control Coeff. | Treatment—control 5% limit | Treatment—control p-Value |
|---|---|---|---|---|---|---|
| *Panel A: Karnataka Wave 1 Households* | | | | | | |
| 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 | 37, 861.564 | 129, 797.423 | 1351.740 | 11, 597.294 | 0.819 |
| Network (friends and family) loans | 7218 | 2735.470 | 25, 394.731 | 6.467 | 1716.401 | 0.994 |
| SHG loans | 7218 | 2543.994 | 6944.324 | 14.783 | 968.668 | 0.976 |
| Bank loans | 7218 | 19, 892.356 | 106, 358.225 | 3563.106 | 8808.589 | 0.428 |
| Moneylender loans | 7218 | 3638.339 | 20, 456.671 | −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* | | | | | | |
| 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 |

(continued)

TABLE 1
*Continued*

| | | | | Treatment—control | | |
|---|---|---|---|---|---|---|
| | Obs | Control mean | Control SD | Coeff. | 5% limit | *p*-Value |
| 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 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 MF. In Hyderabad, 104 neighbourhoods were subject to randomized assignment of MF. 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 neighbourhood level, while "endline 1" values are at the household level. $H(RF)$ is the random forest classification into high ($H=1$) or low ($H=0$) MF propensity. $H(logit)$ is the corresponding logit classification: see Section 2.4 for details.

### 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. Online Appendix Table D.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 Online Appendix Table D.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-MF villages in Wave 2 in Online Appendix Table D.2, Panel A.

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

13. In Auxiliary Appendix Tables C.1 and C.3, 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.

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 D.2, Panel B, we show the means of network characteristics in control neighbourhoods. In this urban sample, the networks are even more sparse than in Karnataka; the average degree is 6.0, for an average neighbourhood 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 MF, we need to identify which households *would have* taken out MF loans in the non-MF villages or neighbourhoods, 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 MF as a function of baseline characteristics, in the presence of MF. 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 MF take-up is based on BSS's rules: only households with a female in the age range of 18–57 were eligible for MF. 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 MF 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 MF 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 MF; (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 MF; (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. The details of the estimation algorithm, implemented choices, and quality are presented in Online Appendix C.

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 MF and non-MF villages (or neighbourhoods) to classify each household as *H* or *L* (high or low likelihood of joining MF).

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.

An advantage of using random forests is that they naturally allow for non-linearities and potentially complex interactions between characteristics that could drive MF take-up.[16] A related advantage of random forest comes from its value in identification. Because 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 use logistic regression to classify households into $H$ and $L$ types for both Karnataka and Hyderabad.[17] In Online Appendix Section C.5, we investigate classification quality from both approaches. We show that random forest performs comparably to logit in Karnataka, while random forest outperforms logit on all metrics in Hyderabad, sometimes quite substantially. We present our main results using the logistic classification in Online Appendix G. The results are quite similar in the Karnataka sample, but are, unsurprisingly, noisier in the Hyderabad 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 MF villages. $H$ types borrow Rs. 1785 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-MF 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 MF entered the treated neighbourhoods, $H$ types had considerably more microcredit than $L$ types in early MF neighbourhoods (Rs. 8773, $p < 0.001$).

## 3. CHANGES IN NETWORKS

How does exposure to MF 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.

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.

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
*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 | 4520 | 4483 | 4516 | 4516 | 4508 |
| 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 | 4523 | 4523 | 4523 | 4523 |
| Depvar Mean | 5.806 | 4.379 | 1.428 | 0.074 |

*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-MF (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.

### 3.1. *Effect on the total number of links*

To look at how introducing MF 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 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 $\text{Post}_t$. 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 MF. 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 *et al*., 2014a, 2014b) to select the controls.[19]

We find that exposure to MF leads to a drop in network density by about 1.2–1.3pp relative to a mean of 11.4% in non-MF 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 MF 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 neighbourhoods 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 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 MF exposure.

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

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 MF in the same regression.

TABLE 3

*Effects of MF 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 |
|---|---|---|---|---|---|---|---|---|---|
| MF × Post | −0.012 | −0.013 | −0.013 | 0.004 | 0.010 | 0.010 | −0.022 | −0.018 | −0.016 |
| | (0.007) | (0.007) | (0.007) | (0.015) | (0.007) | (0.016) | (0.010) | (0.012) | (0.012) |
| | [0.084] | [0.068] | [0.083] | [0.807] | [0.083] | [0.533] | [0.024] | [0.128] | [0.188] |
| Microfinance | −0.020 | 0.006 | 0.007 | −0.041 | −0.018 | −0.013 | −0.013 | 0.011 | 0.012 |
| | (0.008) | (0.005) | (0.006) | (0.016) | (0.006) | (0.015) | (0.010) | (0.010) | (0.011) |
| | [0.018] | [0.277] | [0.235] | [0.013] | [0.235] | [0.377] | [0.199] | [0.239] | [0.258] |
| Post | −0.012 | −0.014 | −0.014 | −0.009 | 0.009 | 0.009 | 0.105 | 0.064 | −0.042 |
| | (0.006) | (0.011) | (0.011) | (0.010) | (0.011) | (0.026) | (0.008) | (0.057) | (0.055) |
| | [0.046] | [0.183] | [0.206] | [0.366] | [0.206] | [0.745] | [0.000] | [0.266] | [0.448] |
| Observations | 150 | 150 | 150 | 150 | 150 | 150 | 150 | 150 | 150 |
| Double-Post LASSO | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Village FE | | | ✓ | | | ✓ | | | ✓ |
| Non-MF Mean | 0.113 | 0.113 | 0.113 | 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 |
|---|---|---|---|---|---|---|
| MF | −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 | 4429 | 4429 | 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 $H$ versus $L$. 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.

### 3.2. *How are links affected by MF?*

In this subsection, we explore how MF exposure affects the formation of links across types of households.

Before turning to results using the predicted $H$s 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 MF takers (*i.e.* $H$s). Note that in Karnataka only 18.5% of households in treatment villages took up MF (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 MF 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 MF *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 MF 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, we can still smoothly control for the features and allow the control to vary by whether the village is exposed to MF or not. This allows us to control for potentially differential effect of MF 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, $\beta_{LH}$ and $\beta_{HH}$, capture whether being in a MF 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 *et al.* (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.

20. In the absence of MF, 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 MF takers for the average village (with 223 households) in the absence of MF, which is 678. The density reduction observed in MF villages corresponds to an overall reduction of 317 total links, which would represent 47% of the total number of counterfactual links involving MF 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.

TABLE 4
*Link evolution, Karnataka*

| | (1) Linked Post-MF | (2) Linked Post-MF | (3) Linked Post-MF | (4) Linked Post-MF |
|---|---|---|---|---|
| MF | −0.058 | −0.059 | −0.023 | −0.021 |
| | (0.018) | (0.019) | (0.008) | (0.008) |
| | [0.002] | [0.002] | [0.006] | [0.008] |
| MF × *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] |
| MF × *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 × *LH* = 0 *p*-val | 0.014 | 0.005 | 0.015 | 0.014 |
| MF + MF × *HH* = 0 *p*-val | 0.361 | 0.088 | 0.101 | 0.232 |
| MF + *LH* × MF = MF + *HH* × 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.

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 does not 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 $\beta$, the coefficient on MF, which captures the effect on the omitted category of links (*LL*), as well as $\beta_{LH}$ and $\beta_{HH}$, which ask whether MF 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 MF villages on a base of 7.5% in non-MF villages ($p = 0.006$ without controls and $p = 0.008$ with controls). Again, the effect is comparable for *LH* links, *i.e.* $\beta_{LH}$ is small and not

TABLE 5
*Link evolution, Hyderabad*

| | (1) Prob. Linked | (2) Prob. Linked |
|---|---|---|
| MF | −0.005 | −0.007 |
| | (0.002) | (0.002) |
| | [0.035] | [0.004] |
| MF × *LH* | 0.002 | −0.001 |
| | (0.003) | (0.002) |
| | [0.577] | [0.764] |
| MF × *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 × *LH* = 0 *p*-val | 0.387 | 0.019 |
| MF + MF × *HH* = 0 *p*-val | 0.066 | 0.041 |
| MF + MF × *HH* = MF + MF × *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.

significant (either with or without controls). The negative effect on link formation is somewhat less pronounced for *HH* links: $\beta_{HH}$ is positive and significant with controls (but not significant without controls.

The differential changes in network structure in the MF villages shed light on network formation. The fact that the *LH* links break may reflect the fact that the *H*s are no longer interested in maintaining their links with the *L*s 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 *L*s should have a stronger incentive to hold on to their mutual links precisely because they no longer have access to the links with the *H*s.[22]

We turn to the Hyderabad data in Table 5. In this case, while MF 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 "MF" 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 neighbourhoods relative to non-microfinance neighbourhoods ($p = 0.035$). 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.

22. In Auxiliary Appendix C, 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.

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 analysed 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 MF 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 MF 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 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 *H*s, 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 MF—are at least as likely to be affected as relationships involving *H*s. 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 *et al.* (2008), Ambrus *et al.* (2014), and Jackson *et al.* (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 Auxiliary Appendix A.2). These interdependencies in link formation could in principle explain the impact of MF 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 MF 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 *L*s should experience minimal impacts.

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 |
|---|---|---|---|---|
| MF | −0.052 | −0.012 | −0.050 | −0.015 |
| | (0.021) | (0.005) | (0.018) | (0.005) |
| | [0.014] | [0.016] | [0.006] | [0.007] |
| MF × *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] |
| MF × *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 × *LH* = 0 *p*-val | 0.005 | 0.035 | 0.008 | 0.025 |
| MF + MF × *HH* = 0 *p*-val | 0.343 | 0.157 | 0.615 | 0.061 |
| MF + *LH* × MF = MF + *HH* × MF *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) |
|---|---|---|---|---|
| MF | −0.363 | −0.393 | −0.195 | −0.157 |
| | (0.137) | (0.140) | (0.118) | (0.113) |
| | [0.010] | [0.006] | [0.101] | [0.169] |
| MF × *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 | 4429 | 4429 | 4429 | 4429 |
| 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 versus 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 MF indicator.

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:

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

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 MF. 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 MF relative to non-MF villages ($p = 0.008$). The positive coefficients on the terms $\beta_{LLH}, \beta_{LHH}$ and $\beta_{HHH}$ 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 MF 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 1000 for readability.[23] Although the results are noisier than in Karnataka, we find once again that *LLL* triangles are negatively affected by MF: in column 1, we are 50% less likely to see any *LLL* triangle in MF neighbourhoods ($p = 0.067$). 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 MF. 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 Auxiliary Appendix C, 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.

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.

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 |
|---|---|---|---|---|
| MF | −0.078 | −0.070 | −0.085 | −0.076 |
| | (0.029) | (0.026) | (0.023) | (0.019) |
| | [0.008] | [0.008] | [0.000] | [0.000] |
| MF × *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] |
| MF × *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] |
| MF × *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 × *HHH* = 0 *p*-val | 0.698 | 0.549 | 0.935 | 0.209 |
| MF + MF × *LLH* = 0 *p*-val | 0.023 | 0.03 | 0.022 | 0.025 |
| MF + MF × *LHH* = 0 *p*-val | 0.262 | 0.048 | 0.141 | 0.018 |
| MF + MF × *HHH* = MF + MF × *LLH* *p*-val | 0.076 | 0.35 | 0.093 | 0.459 |
| MF + MF × *HHH* = MF + MF × *LHH* *p*-val | 0.212 | 0.492 | 0.075 | 0.307 |
| MF + MF × *LLH* = MF + MF × *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.

## 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 E). This is in contrast with local externalities where the value of a link depends only on the other connections of the people

TABLE 8
*Triples evolution, Hyderabad*

| All variables × 1000 | Full Triangle Linked (1) | Full Triangle Linked (2) |
|---|---|---|
| MF | −0.018 | −0.034 |
| | (0.010) | (0.020) |
| | [0.067] | [0.086] |
| MF × *LLH* | 0.010 | −0.012 |
| | (0.011) | (0.013) |
| | [0.370] | [0.344] |
| MF × *LHH* | −0.027 | −0.052 |
| | (0.038) | (0.040) |
| | [0.472] | [0.191] |
| MF × *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.

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 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 MF, we describe it in a general form and then specialize to the two-type $(H, L)$ MF case.

## 4.1. *Types and utilities*

There are $n$ individuals, indexed by $i, j \cdots \in \{1, \ldots, n\}$. Agent $i$ has a type $\theta_i$ from a type set $\Theta$. Let $v_{\theta\theta'}$ denote the base benefit that an agent of type $\theta$ gets from a relationship with an agent of type $\theta'$. 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 $\varepsilon_{ij}$ 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 $\varepsilon_{ij}$ is distributed according to an atomless distribution $F$.

A useful expression is

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

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 $\theta$ then has an expected utility from $d_{\theta\theta'}$ friends of type $\theta'$ 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 $e_ie_j$ 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 agents $i$ and $j$ keep their friendship with probability $e_ie_j$ and lose it with probability $1-e_ie_j$. If $g_{ij}^{t-1}=0$, then agents $i$, $j$ were not connected in the last period and form a friendship with probability

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

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.

24. This is a useful and conventional modeling device. See Currarini *et al*. (2009), Currarini *et al*. (2010), Cabrales *et al*. (2011) and Canen *et al*. (2017) for other models where socialization takes effort and there is random meeting.

### 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 $e_\theta$ 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 $\theta'$ with whom an agent of type $\theta$ could potentially form friendships. If $\theta' \neq \theta$, then this is generally the number of agents of type $\theta'$,[26] while if it is of type $\theta$, 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 $\theta$, 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 $\theta'$ with which an agent of type $\theta$ 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.

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

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

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 $v$s.

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

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

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

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

### 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 $\theta$, 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 $\theta$. 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 $\theta$), and agents of the same type take the same efforts. In addition, if $\mathrm{E}^+[\upsilon_{\theta\theta'}] > 0, n_{\theta\theta'} > 0$ for each $\theta, \theta'$,*[30] *and $\upsilon_{\theta\theta'}$ is decreased for some $\theta\theta'$ (holding all other parameters constant), then $e_{\theta''}$ decreases for all $\theta''$, 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 $\theta, \theta'$ entries

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

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 $\upsilon_{\theta\theta'}$ 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.

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 $\theta$ and $\theta'$ that include all types. Note also that $\mathrm{E}^+[\upsilon_{\theta\theta'}] > 0$ 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.

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 MF). The intuition is that if marginal benefits to $L$s from connecting to $H$s are particularly high, then when $H$s drop effort, payoffs from effort for $L$s can drop even more than for $H$s, leading to an even larger effect on $L$s. See Online Appendix I 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 MF*

We next present a specialized case of the model to analyse how MF 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 MF maps to the model parameters, specifically the $\upsilon_{\theta,\theta'}$s.

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

$$\begin{aligned}
c_H e_H^\star &= u_H + \mathrm{E}^+[\upsilon_{HH}]\,(\lambda n - 1)\,(1 - F(-\upsilon_{HH}))^2\, e_H^\star \\
&\quad + \mathrm{E}^+[\upsilon_{HL}]\,(1-\lambda)\,n\,(1 - F(-\upsilon_{HL}))\,(1 - F(-\upsilon_{LH}))\, e_L^\star, \\
c_L e_L^\star &= u_L + \mathrm{E}^+[\upsilon_{LL}]\,((1-\lambda)\,n - 1)\,(1 - F(-\upsilon_{LL}))^2\, e_L^\star \\
&\quad + \mathrm{E}^+[\upsilon_{LH}]\lambda n\,(1 - F(-\upsilon_{LH}))\,(1 - F(-\upsilon_{HL}))\, e_H^\star, \\
d_{HL} &= ((1-\lambda)\,n)\, e_H^\star e_L^\star\,(1 - F(-\upsilon_{HL}))\,(1 - F(-\upsilon_{LH}))\,, \\
d_{LH} &= d_{HL}\frac{\lambda}{1-\lambda}, \\
d_{HH} &= (\lambda n - 1)\,(e_H^\star)^2\,(1 - F(-\upsilon_{HH}))^2\,, \\
d_{LL} &= ((1-\lambda)n - 1)\,(e_L^\star)^2\,(1 - F(-\upsilon_{LL}))^2\,.
\end{aligned}$$

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 MF: an example.** How does the entry of MF affect these parameters? Here, we present a rationalization for the payoffs based on informal borrowing and lending. In

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

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 $H$s 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 $\alpha_H$ of having money to lend and a 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 $\alpha_L$ and $\beta_L$ 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 MF changes these parameters. There are several likely channels by which the introduction of MF 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 $\beta_H$ 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 $\alpha_H$ and perhaps an increase in $\beta_H$. 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] $\beta_H$ 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 MF. Her probability of needing to borrow may also go down to $\beta'_H \leq \beta_H$ if MF loans are a substitutes for network credit. In any case, we maintain that the $L$s' needs for borrowing and lending are unaltered by the introduction of MF. 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 MF. 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

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

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.

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).

As we describe above, the arrival of MF 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 $\Delta_{HL}$ should be positive whereas $\Delta_{LH}$ 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 $\alpha_H$ and $\beta_H$ go down. In this case, as long as both $b$ and $r$ are positive, all of $\upsilon_{HL}$, $\upsilon_{LH}$ and $\upsilon_{HH}$ 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 I, we use simulations to show that it is indeed possible that *LL* links may be most affected, given the decreased effort of *H*s as well as the *H*s' lower propensity to want to link with *L*s given their decreased borrowing needs. Specifically, we consider the case where only $\upsilon_{HL}$ declines when MF enters and look at the resulting declines in efforts as a function of the pre-MF levels of $\upsilon_{HL}$ and $\upsilon_{LH}$. 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 *L*s due to the entry of MF causes *L*s to reduce socialization even more than *H*s. This is because in equilibrium, mutual consent with the *H*s 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 in Online Appendix B.

### 4.8. *Alternative explanations*

In this section we 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 account for the facts in combination with a simpler model of network formation?

**4.8.1. Alternative models of network formation.** In Online Appendix E, 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.

35. See the calculations in Online Appendix J.

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[36] 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 *et al.*, 2009; Jackson *et al.*, 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 *et al.*, 2010; König *et al.*, 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 *et al.*, 2009, 2010; Cabrales *et al.*, 2011; Canen *et al.*, 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 a model of undirected search, and makes it difficult to explain why *LL* and *LLL* relationships drop in response to a decrease in *H*s' willingness to link to *L*s. Our model relies on the idea that individuals put effort into trying to meet but cannot choose who they are meeting. The reason *H*s lower effort less than *L*s is that they are (correctly) more optimistic about actually linking with those who they meet. If on the other hand, the *L*s 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 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.[37] 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 Online Appendix B, 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 *H*s 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

36. 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.

37. The models by Currarini *et al.* (2009, 2010) and Canen *et al.* (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 MF is introduced.

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 simultaneously; 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 Online Appendix E, we discuss why dropping any one of these features would fail to capture some aspects of the data.

**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 *H*s, she gives up the entire project and therefore also stops borrowing from other *L*s. 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 E.3.2 for more details.

Another possibility is that the reason *LL* links drop is that *L*s recognize that even if they do not participate in MF, it is available to them. This is probably true for some of them, but because we use MF 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 *H*s tend to socialize with *H*s in MF meetings would provide a force unique to participants, hence *H*s, 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 Auxiliary Appendix B. We show that our main results hold even if we condition on all pairs where neither member joined MF (86% of baseline links). Even under this restriction, the *H* classification has content – *H*s are more suitable for MF by construction, and therefore even non-borrowers have higher option value from future access to MF.[38] However, we acknowledge that these results are only suggestive given that actual MF 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 *L*s anymore. Notice our general form of the model allows for this.

38. This option value may in turn reduce the value non-borrowing *H*s receive from maintaining and forming network relationships.

A final possibility is that the entry of MF 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 MF, which finds little impact of MF entry on average village or neighbourhood level outcomes such as consumption, investment or business profit (Angelucci *et al*., 2015; Attanasio *et al*., 2015; Augsburg *et al*., 2015; Banerjee *et al*., 2015b, 2015a; Crépon *et al*., 2015; Tarozzi *et al*., 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 MF 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 MF. In both the Karnataka and Hyderabad data we have rich borrowing information, and we can measure the impacts of MF on several different types of household borrowing (*e.g.* MF, 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 MF village, post, and household type in the Karnataka sample:

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

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 $H$s borrow much more than $L$s do (by 1787 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 MF.[39,40]

Panel B of Table 9 turns to the Hyderabad data. Here we present impacts on borrowing measured in the first post-MF survey wave, 15–18 months after MF was introduced in treatment neighbourhoods.[41]

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

39. SHGs are groups of women who get together to lend to each other.

40. What is striking is that even $L$s 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 E.1 in Auxiliary Appendix E.

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

TABLE 9
*Borrowing patterns*

| | (1) MFI | (2) Friends | (3) SHG | (4) Moneylender | (5) Family |
|---|---|---|---|---|---|
| *Panel A: Borrowing Patterns, Karnataka* | | | | | |
| MF × 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] |
| MF × Post × $H$ | 1795.233 | 203.926 | 48.466 | −2210.964 | −1608.814 |
| | (245.414) | (242.383) | (346.884) | (943.562) | (1185.489) |
| | [0.000] | [0.401] | [0.889] | [0.020] | [0.175] |
| MF × $H$ | −0.542 | −65.457 | 232.443 | 206.495 | 1088.834 |
| | (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 | 1828.811 | 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 × Post × $H$ + MF × Post = 0 $p$-val | 0.000 | 0.255 | 0.119 | 0.084 | 0.325 |
| *Panel B: Borrowing Patterns, Hyderabad* | | | | | |
| MF | −209.748 | 86.742 | −1882.840 | −2664.192 | −256.318 |
| | (235.127) | (894.331) | (801.110) | (1455.603) | (656.431) |
| | [0.375] | [0.923] | [0.021] | [0.071] | [0.697] |
| MF × $H$ | 8312.670 | −637.232 | −1577.128 | 4689.554 | 1796.860 |
| | (448.982) | (1491.449) | (1369.064) | (2622.331) | (1366.622) |
| | [0.000] | [0.671] | [0.252] | [0.077] | [0.192] |
| $H$ | −108.232 | −1792.590 | 1251.211 | −198.590 | −507.290 |
| | (296.017) | (1293.944) | (1163.829) | (1899.306) | (985.862) |
| | [0.716] | [0.169] | [0.285] | [0.917] | [0.608] |
| Observations | 6811 | 6863 | 6863 | 6863 | 6863 |
| 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 MF 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 MF 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 $H$ versus $L$) 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.

Starting with column 1, we find no impact of MF on borrowing for $L$s but a large and significantly greater impact for $H$ types (8776 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 (−1883, $p = 0.021$) and moneylenders (−2664, $p = 0.071$). Note that $H$ types experience no decline in moneylender borrowing with access to MF. In fact they gain Rs. 4690 ($p = 0.077$) 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 MF has an adverse effect on the network borrowing of the $L$s in both contexts. This is especially striking because, all else equal, we would have expected *L*s' informal borrowing to go down less than that of the *H*s, or even to go up to the extent there is re-lending of MF loans (from *H*s to *L*s).

### 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 MF 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,[42] 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. MF (*i.e.* treated) villages are denoted with *MF*. The term $\alpha_i$ is a household-fixed effect which, as well as controlling 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 $\gamma_{vt}$ is a area-time fixed effect which captures the aggregate shock to a particular neighbourhood in a given survey wave. We estimate the following "long differences" specification, using the first and third waves of the data,[43]

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

The coefficient $\beta_1$ 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 $\beta_2$. The coefficient $\beta_3$ captures the extent of income pass-through for $H$ households in control areas. Finally $\beta_4$ measures any differential treatment effect on risk sharing for $H$ households.[44]

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,

42. Recall, the Hyderabad data only contain network information collected in 2012, but income and consumption were collected in 2007–8, 2010 and 2012.

43. 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 MF, 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).

44. In Online Appendix Table H.1, we show that MF does not have any impact on income for $L$ types. While the point estimates are noisy, the coefficient estimate is larger and closer to standard significance levels for the $H$ types.

TABLE 10
*Risk sharing, Hyderabad*

| | (1) Expenditures: Non-Food | (2) Expenditures: Total |
|---|---|---|
| MF × Income | 0.071 | 0.066 |
| | (0.030) | (0.037) |
| | [0.022] | [0.079] |
| MF × 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 × Income + MF × Income × 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-neighbourhood-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 neighbourhood level) are reported in parentheses. $p$-values are reported in brackets.

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 MF 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, MF causes no detectable change to income pass-through ($\beta_2 + \beta_4$, $p = 0.834$).

While inspired by the reduction in links for $L$s, 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 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.[45]

In summary, these results demonstrate that, while $H$ households' ability to buffer income risk is unaffected by MF exposure (or even weakly improved), this is not true for $L$ households.

45. Angelucci and De Giorgi (2009), for example, analyse 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 analyse the interplay between policies and risk sharing, highlighting the potential for risk sharing to go down due to improved outside options.

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 MF 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 MF 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 MF 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.

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.

## 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[46]

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

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

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

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

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

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 $\theta, \theta'$ entries

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

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.

*Acknowledgments.* We thank Siwan Anderson, Patrick Francois, Rachel Kranton, and participants at various seminars for suggestions. We thank Devika Lakhote, Anirudh Sankar, Varun Kapoor, Stephen Nei, Mounu Prem, Tristan Loisel, and Sumhith Veda Aradhyula as well as the CMF at IFMR, Tanay Balantrapu, Gowri Nagraj, and Manaswini Rao for excellent research assistance and helpful discussions.

**Supplementary Data**

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

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

**Funding**

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

**Data and Code Availability Statement**

Code and data (all collected by the authors) used for analysis in this article are available with full documentation in the Zenodo online repository for *ReStud* at: https://doi.org/10.5281/zenodo.7706650 (Banerjee *et al*., 2023).

## REFERENCES

ALBARRAN, P. and ATTANASIO, O. P. (2003), "Limited Commitment and Crowding out of Private Transfers: Evidence from a Randomised Experiment", *The Economic Journal*, **113**, C77–C85.

ALTONJI, J. G., ELDER, T. E. and TABER, C. R. (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, **113**, 151–184.

AMBRUS, A., MOBIUS, M. and SZEIDL, A. (2014), "Consumption Risk-Sharing in Social Networks", *American Economic Review*, **104**, 149–182.

ANGELUCCI, M. and DE GIORGI, G. (2009), "Indirect Effects of an Aid Program: How do Cash Transfers Affect Ineligibles' Consumption?", *American Economic Review*, **99**, 486–508.

ANGELUCCI, M., DE GIORGI, G. and RASUL, I. (2018), "Consumption and Investment in Resource Pooling Family Networks", *The Economic Journal*, **128**, 2613–2651.

ANGELUCCI, M., KARLAN, D. and ZINMAN, J. (2015), "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco", *American Economic Journal: Applied Economics*, **7**, 151–182.

ARROW, K. J. (2000), "Observations on Social Capital", *Social Capital: A Multifaceted Perspective*, **6**, 3–5.

ATTANASIO, O., AUGSBURG, B., DE HAAS, R., *et al.* (2015), "The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia", *American Economic Journal: Applied Economics*, **7**, 90–122.

AUGSBURG, B., DE HAAS, R., HARMGART, H., *et al.* (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 GOYAL, S. (2000), "A Noncooperative Model of Network Formation", *Econometrica*, **68**, 1181–1229.

BANERJEE, A., BREZA, E., CHANDRASEKHAR, A. G., *et al.* (2023), "Data for: Changes in Social Network Structure in Response to Exposure to Formal Credit Markets", *Review of Economic Studies*, forthcoming. https://doi.org/10.1126/science.1236498.

BANERJEE, A., BREZA, E., DUFLO, E., *et al.* (2019a), "Can Microfinance Unlock a Poverty Trap for Some Entrepreneurs?" (Technical report, National Bureau of Economic Research).

BANERJEE, A., CHANDRASEKHAR, A., DUFLO, E., *et al.* (2013), "Diffusion of Microfinance", *Science*, **341**, 363–370. https://doi.org/10.5281/zenodo.7706650.

BANERJEE, A., CHANDRASEKHAR, A. G., DUFLO, E., *et al.* (2019b), "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials", *The Review of Economic Studies*, **86**, 2453–2490.

BANERJEE, A., DUFLO, E., GLENNERSTER, R., *et al.* (2015a), "The Miracle of Microfinance? Evidence from a Randomized Evaluation", *American Economic Journal: Applied Economics*, **7**, 22–53.

BANERJEE, A., KARLAN, D. and ZINMAN, J. (2015b), "Six Randomized Evaluations of Microcredit: Introduction and Further Steps", *American Economic Journal: Applied Economics*, **7**, 1–21.

BEAMAN, L. and MAGRUDER, J. (2012), "Who Gets the Job Referral? Evidence from a Social Networks Experiment", *American Economic Review*, **102**, 3574–3593.

BELLONI, A. and CHERNOZHUKOV, V. (2009), "Least Squares After Model Selection in High-Dimensional Sparse Models" (Working Paper, MIT Department of Economics).

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014a), "High-Dimensional Methods and Inference on Structural and Treatment Effects", *The Journal of Economic Perspectives*, **28**, 29–50.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014b), "Inference on Treatment Effects after Selection Among High-Dimensional Controls", *The Review of Economic Studies*, **81**, 608–650.

BINZEL, C., FIELD, E. and PANDE, R. (2013), "Does the Arrival of a Formal Financial Institution Alter Informal Sharing Arrangements? Experimental Evidence from Village India" (Working Paper).

BLOCH, F., GENICOT, G. and RAY, D. (2008), "Informal Insurance in Social Networks", *Journal of Economic Theory*, **143**, 36–58.

BLUMENSTOCK, J. E., EAGLE, N. and FAFCHAMPS, M. (2016), "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters", *Journal of Development Economics*, **120**, 157–181.

BLUMENSTOCK, J. and TAN, X. (2016), "Social Networks and Migration: Theory and Evidence from Rwanda" (Working Paper).

BOUCHER, V. (2015), "Structural Homophily", *International Economic Review*, **56**, 235–264.

BREIMAN, L. and SPECTOR, P. (1992), "Submodel Selection and Evaluation in Regression. The X-Random Case", *International Statistical Review/Revue Internationale de Statistique*, **60**, 291–319.

BREZA, E. (2016), “Field Experiments, Social Networks, and Development” (The Oxford Handbook of the Economics of Networks).
BREZA, E. and CHANDRASEKHAR, A. G. (2019), “Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment”, *Econometrica*, **87**, 175–216.
BREZA, E., CHANDRASEKHAR, A. G., MCCORMICK, T. H., *et al.* (2019), “Consistently Estimating Graph Statistics Using Aggregated Relational Data” (arXiv preprint arXiv:1908.09881).
BREZA, E., CHANDRASEKHAR, A. G., MCCORMICK, T. H., *et al.* (2020), “Using Aggregated Relational Data to Feasibly Identify Network Structure Without Network Data”, *American Economic Review*, **110**, 2454–2484.
BREZA, E. and KINNAN, C. (2021), “Measuring the Equilibrium Impacts of Credit: Evidence from the Indian Microfinance Crisis”, *The Quarterly Journal of Economics*, **136**, 1447–1497.
CABRALES, A., CALVÓ-ARMENGOL, A. and ZENOU, Y. (2011), “Social Interactions and Spillovers”, *Games and Economic Behavior*, **72**, 339–360.
CANEN, N., JACKSON, M. O. and TREBBI, F. (2017), “Endogenous Networks and Legislative Activity” (SSRN http://ssrn.com/abstract=2823338).
CHANDRASEKHAR, A. and LEWIS, R. (2014), “Econometrics of Sampled Networks” (Working Paper, MIT).
CHRISTAKIS, N., FOWLER, J., IMBENS, G., *et al.* (2010), “An Empirical Model for Strategic Network Formation” (Working Paper, NBER).
COATE, S. and RAVALLION, M. (1993), “Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements”, *Journal of Development Economics*, **40**, 1–24.
COMOLA, M. and PRINA, S. (2021), “Treatment Effect Accounting for Network Changes”, *Review of Economics and Statistics*, **103**, 1–25.
CRÉPON, B., DEVOTO, F., DUFLO, E., *et al.* (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., JACKSON, M. O. and PIN, P. (2009), “An Economic Model of Friendship: Homophily, Minorities, and Segregation”, *Econometrica*, **77**, 1003–1045.
CURRARINI, S., JACKSON, M. O. and PIN, P. (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 MORELLI, M. (2000), “Network Formation with Sequential Demands”, *Review of Economic Design*, **5**, 229–249.
DUPAS, P., KEATS, A. and ROBINSON, J. (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 MUTUSWAMI, S. (1997), “Stable Networks”, *Journal of Economic Theory*, **76**, 322–344.
FAFCHAMPS, M. and LUND, S. (2003), “Risk-sharing Networks in Rural Philippines”, *Journal of Development Economics*, **71**, 261–287.
FEIGENBERG, B., FIELD, E. and PANDE, R. (2013), “The Economic Returns to Social Interaction: Experimental Evidence from Microfinance”, *Review of Economic Studies*, **80**, 1459–1483.
FIELD, E., PANDE, R., PAPP, J., *et al.* (2012), “Repayment Flexibility can Reduce Financial Stress: A Randomized Control Trial with Microfinance Clients in India”, *PLoS ONE*, **7**, e45679.
HAYASHI, F., ALTONJI, J. and KOTLIKOFF, L. (1996), “Risk-Sharing Between and Within Families”, *Econometrica: Journal of the Econometric Society*, **64**, 261–294.
HERINGS, P. J.-J., MAULEON, A. and VANNETELBOSCH, V. (2009), “Farsightedly Stable Networks”, *Games and Economic Behavior*, **67**, 526–541.
HESS, S. H., JAIMOVICH, D. and SCHÜNDELN, M. (2020), “Development Projects and Economic Networks: Lessons From Rural Gambia”, *The Review of Economic Studies*, **88**, 1347–1384.
HOFF, P., RAFTERY, A. and HANDCOCK, M. (2002), “Latent Space Approaches to Social Network Analysis”, *Journal of the American Statistical Association*, **97**, 1090–1098.
ISLAM, A., NGUYEN, C. and SMYTH, R. (2015), “Does Microfinance Change Informal Lending in Village Economies? Evidence from Bangladesh”, *Journal of Banking & Finance*, **50**, 141–156.
JACKSON, M. O. (2003), “The Stability and Efficiency of Economic and Social Networks”, in Koray, S. and Sertel, M (eds) *Advances in Economic Design* (Heidelberg: Springer-Verlag).
JACKSON, M. O. (2008), *Social and Economic Networks* (Princeton: Princeton University Press).
JACKSON, M. O., RODRIGUEZ-BARRAQUER, T. R. and TAN, X. (2012), “Social Capital and Social Quilts: Network Patterns of Favor Exchange”, *American Economic Review*, **102**, 1857–1897.
JACKSON, M. O. and VAN DEN NOUWELAND, A. (2005), “Strongly Stable Networks”, *Games and Economic Behavior*, **51**, 420–444.
JACKSON, M. O. and WATTS, A. (2002), “The Evolution of Social and Economic Networks”, *Journal of Economic Theory*, **106**, 265–295.
JACKSON, M. O. and WOLINSKY, A. (1996), “A Strategic Model of Social and Economic Networks”, *Journal of Economic Theory*, **71**, 44–74.
KARLAN, D., MOBIUS, M., ROSENBLAT, T., *et al.* (2009), “Trust and Social Collateral”, *The Quarterly Journal of Economics*, **124**, 1307–1361.
KINNAN, C. and TOWNSEND, R. (2012), “Kinship and Financial Networks, Formal Financial Access, and Risk Reduction”, *American Economic Review*, **102**, 289–293.

KÖNIG, M. D., TESSONE, C. J. and ZENOU, Y. (2014), "Nestedness in Networks: A Theoretical Model and Some Applications", *Theoretical Economics*, **9**, 695–752.

KRANTON, R. E. (1996), "Reciprocal Exchange: A Self-Sustaining System", *The American Economic Review*, **86**, 830–851.

LIGON, E., THOMAS, J. P. and WORRALL, T. (2000), "Mutual Insurance, Individual Savings, and Limited Commitment", *Review of Economic Dynamics*, **3**, 216–246.

MACCHIAVELLO, R. and MORJARIA, A. (2021), "Competition and Relational Contracts in the Rwanda Coffee Chain", *The Quarterly Journal of Economics*, **136**, 1089–1143.

MCMILLAN, J. and WOODRUFF, C. (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 ROSENZWEIG, M. (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., DESAI, J. and JOHNSON, K. (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 VIVES, X. (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 GOODACRE, R. (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.