Appendix B. Random Forest Model Description

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

B.1. Algorithm Inputs.

Input Data:

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

Algorithm Parameters:

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

B.2. Basic Algorithm.

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

Step 2: For each tree $t \in T$,

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

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

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

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

B.3. Our Parameter Choices.

- T: We use 1500 trees.
- p: We use 15 predictors and the choice of predictors is explained in subsection B.4.
- m: We use the basic R `randomForest` parameter which is equal to $\sqrt{p}$ for classification.

- c: We use the vector $(0.85, 0.15)$ for Karnataka panel and $(0.73, 0.27)$ for Hyderabad panel, chosen by cross-validation.
- t: We use 0.7 of our sample to train the data.

B.4. **Selection of predictors.** To select our set of predictors, we use the eligbility rules of the microfinance firm as well as network and social distance to leaders in the village, which should predict how likely it is that a household hears of microfinance. These predictors are as follows:

- dummy for being a BSS leader, who are the people that the MFI would approach when entering a village,
- dummy for if the household has a female of eligible age for a microfinance loan, which is a requirement for the household to be able to participate,
- the average closeness (mean of inverse of network distance) to leaders, which is relevant because as in Banerjee et al. (2013) those who are closer to leaders should be more likely to hear of microfinance,
- the average closeness (mean of inverse distance) to same-caste leaders, because interactions within-caste are more likely and therefore should influence the likelihood of being informed,
- the share of same-caste leaders in the village, as above.

We also use these other variables, some of which are related to the five previous ones. Here are a few examples:

- GMOBC= a dummy for whether the household consists of general caste or other backward caste, so the omitted categories are scheduled caste and scheduled tribes (general and OBC are considered upper caste),
- household size,
- number of rooms,
- number of beds,
- dummy for access to electricity,
- dummy for access to latrine,
- dummy for RCC roof,
- dummy for Thatched roof.

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

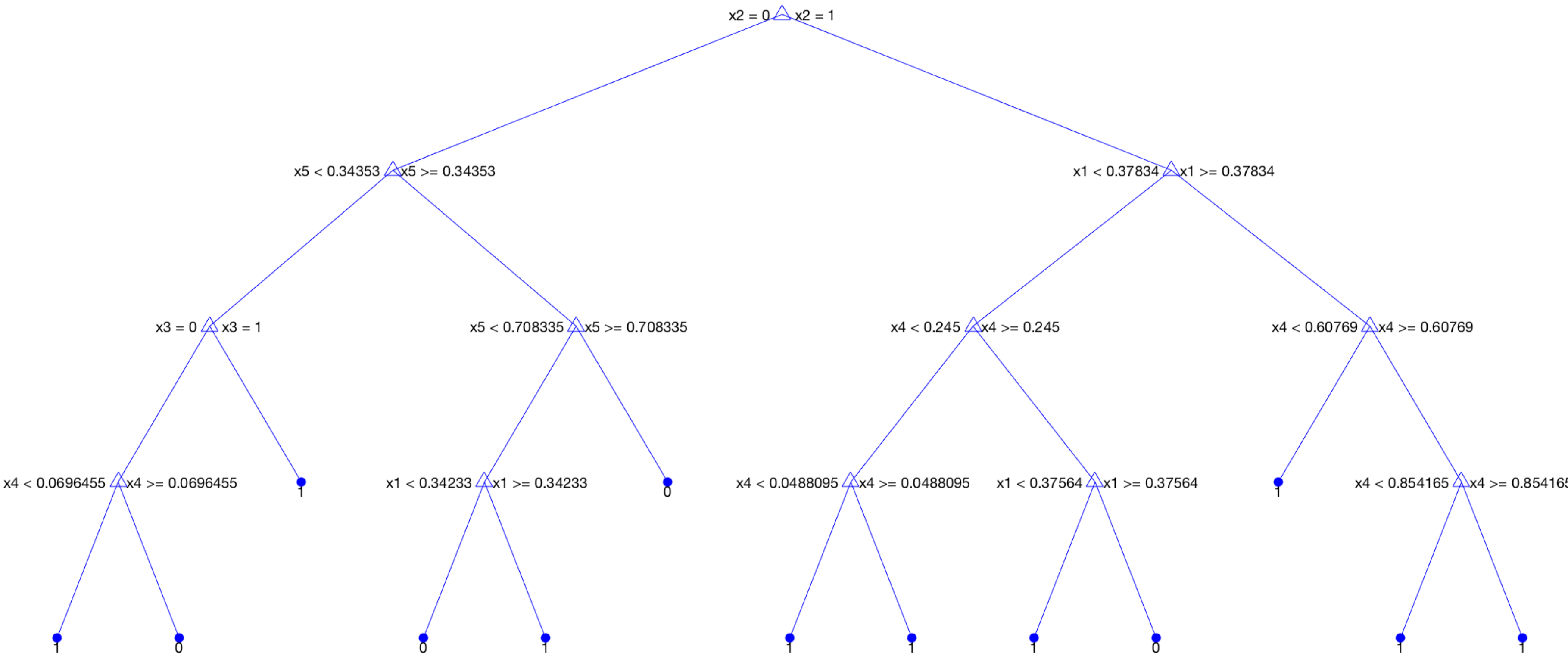


FIGURE B.1. This presents an example of a decision tree. For the sake of simplicity, we limit the maximum number of splits to 12. The actual procedure has a considerably more complex tree. Here x_1 is the average closeness to leaders, x_2 is whether the household is eligible by having a female of eligible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

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

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

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

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

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

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

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

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

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

Both random forest and logistic classifiers have a cut-off parameter that was chosen by 3-fold cross-validation to maximize the Matthews correlation coefficient. We next compare each classification metric across random forest and logistic. In every case, random forest outperforms logit.

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

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

- We interpret RF to have a strong correlation, while logit's is weak.

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

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

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

(1) Confusion Matrix

(a) True negative rate:
- Random forest: 79%
- Logit: 78%

(b) True positive rate:
- Random forest: 45%
- Logit: 35%

(c) Positive predicted value (probability of true positive vs all positive):
- Random forest: 38%
- Logit: 31%

(d) Negative predicted value (probability of true negative vs all negative):
- Random forest: 84%
- Logit: 81%

(2) Metrics for quality of classification are:

(a) Matthews correlation coefficient (from [-1,1]): the preferred diagnostic measure
- Random forest: 0.226
- Logit: 0.120
- We interpret RF to have a strong correlation, while logit's is weak.

(b) F1 score (from $[0, 1]$)
- Random forest: 0.409
- Logit: 0.325

(c) Diagnostic odds ratio (positive odds ratio / negative odds ratio, from $[0, \infty)$):
- Random forest: 3.09
- Logit: 1.87

Appendix C. Balance

In both Hyderabad and Karnataka, the treated (i.e., exposed to microfinance) and control (not exposed to microfinance) samples are balanced in terms of predetermined characteristics. These are the characteristics we use to classify households into those that have a high probability of taking up microfinance loans (H) and those that have a low probability (L), when offered.

Table C.1. Covariate balance

Panel A: Karnataka

				Treatment - control		
	Obs	Control Mean	Control SD	Coeff.	5% limit	p-value
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
WC harm mean dist	7511	0.495	0.186	-0.020	0.028	0.127
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

Panel B: Hyderabad

				Treatment - control		
	Obs	Control Mean	Control SD	Coeff.	5% limit	p-value
Total number of businesses in area, baseline	105	7.288	5.003	-0.346	1.927	0.726
Area mean monthly per-capita exp, baseline	105	1004.974	171.510	42.847	70.733	0.238
Area literacy rate (HH heads), baseline	105	0.625	0.167	0.007	0.056	0.811
Area literacy rate (all), baseline	105	0.687	0.094	0.000	0.032	0.976
HH had a business pre-intervention	6865	0.308	0.462	0.007	0.042	0.736
HH size (adult equiv), endline 1	6865	4.690	1.784	-0.008	0.130	0.899
Adults (16+) in HH, endline 1	6865	3.887	1.754	-0.018	0.129	0.780
Children (<15) in HH, endline 1	6865	1.738	1.310	-0.014	0.104	0.797
Male head of household, endline 1	6865	0.895	0.307	0.012	0.021	0.266
Age of head of household, endline 1	6865	41.146	10.228	-0.226	0.774	0.566
Head of HH with no education, endline 1	6865	0.312	0.463	0.001	0.044	0.975
Any child 13-18 in HH, endline 1	6865	0.452	0.498	0.016	0.031	0.305
Spouse is literate, endline 1	6865	0.543	0.498	0.003	0.049	0.919
Spouse works for a wage, endline 1	6865	0.234	0.423	-0.020	0.048	0.405
H	6865	0.241	0.427	-0.016	0.077	0.684

Notes: This table reports Treatment-Control balance on additional variables used as predictors to classify households as H (high MF propensity) vs. L (low MF propensity). Unit of observation: household (except for area variables). p-values of differences reflect standard errors clustered at the area level.

Table C.2. Endline network summary statistics

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

Appendix D. Alternative Models: Existing Models in the Literature

In this section we describe several alternative models, emphasizing why they cannot generate the patterns in our data; and we also describe an extension of our model that includes direct payoff externalities across links. The goal is to provide the reader with a perspective as to why our model is new and why insights from existing models are insufficient. Of course, this is not an exhaustive list of all models in the literature, but it is representative of the types of models that would be natural candidates for this application.

In what follows, we take the setup of Section 4.6 to work through these models. We study four specific alternatives.

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

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

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

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

The fourth model, presented in Section D.4, returns to the setup where networks essentially re-form every period, but now introduces "directed search". With directed search, agents are free to choose which other types of agents they want to link with. In such a model, we find that while HL links should decline in microfinance villages, LL links should certainly increase. This fits with the main strand of the network formation literature (e.g., Jackson and Wolinsky (1996); Dutta and Mutuswami (1997); Bala and Goyal (2000); Currarini and Morelli (2000); Jackson and Van den Nouweland (2005); Bloch, Genicot, and Ray (2008); Herings, Mauleon, and Vannetelbosch (2009); Jackson, Rodriguez-Barraquer, and Tan (2012); Boucher (2015)...).

Taken together, these four perspectives, which take either exogenous or directed search, with mutual consent and possibly support, cannot generate patterns consistent with the data.

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

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

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

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

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

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

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

compared to

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

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

$$\begin{aligned} G'(v_{HL})\Delta v_{HL} + G'(v_{LH})\Delta v_{LH} &= G'(v_{HL})[\Delta v_{HL} + \Delta v_{LH}] \\ &= (\alpha_H \Delta\beta_H + \beta_H \Delta\alpha_H)(r + b) < 0. \end{aligned}$$

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

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

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

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

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

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

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

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

D.3. **A model with supported links.** Our third model again takes the perspective that links are easy to break but slow to form, but in this case we focus on the value of a link being supported in the sense of Jackson et al. (2012). Jackson, Rodriguez-Barraquer, and Tan (2012) introduce the notion of support, which correlates the presence of links based on incentives to exchange favors (including lending to each other).The idea is that two households in isolation may not have enough bilateral interaction to be able to sustain cooperation with each other, but if they both also have relationships with some other households in common, then the relationships can all "support" each other: if someone fails to cooperate with one of their friends then beyond losing that relationship, they also lose relationships with all the other friends that they had in common with the friend with whom they did not cooperate. Fear of losing all of those relationships if they misbehave provides added incentives to maintain cooperation.[45] This leads relationships to be correlated: forming them in supported combinations provides stronger incentives, and then both their presence and disappearance ends up being correlated.

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

D.3.1. *Payoffs.* We start with a set of HH, LH, and LL links. However some of these links also support each other in the sense that some are part of HHH, LHH, LLH, or LLL triangles. We assume that no one has more than two links, to keep the problem manageable. We assume

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

that the payoff to i from the links between i (a type θ) and j (a type $\widetilde{\theta}$) that is supported by k (a type θ') is given by

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

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

This formulation makes sense in a world where there is no crowd-out in borrowing or lending – when an agent is in the borrowing state he gets twice the benefit b if he can borrow from two sources and when he is in the lending state he gets twice the benefit r if he can lend to two people. The modeling of the relationship specific utility term captures the intuition that when three people are hanging out together, the effect of the relationship to each of them depends on its best parts from their point of view.

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

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

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

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

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

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

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

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

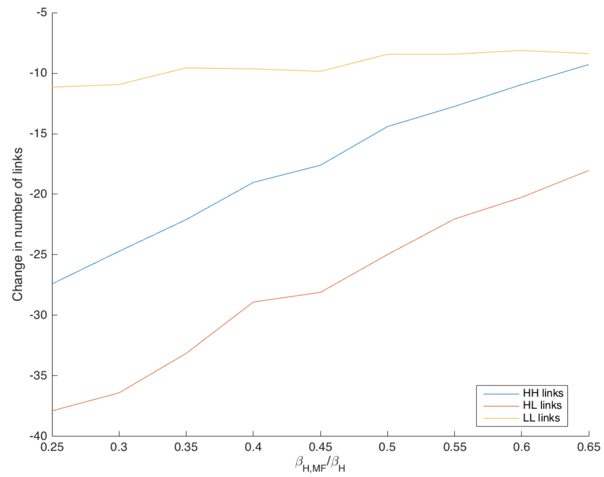
D.3.3. *Simulation.* To make this transparent, we present a simulation exercise. We look at networks of size $n = 300$. We set the payoff parameters $r = 0.1$, $b = 1$, and $\alpha_H = \alpha_L = \beta_H = \beta_L = 1/3$. We set $\alpha'_H = 1.45\alpha_H$ and we will vary the needing to borrow probability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

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

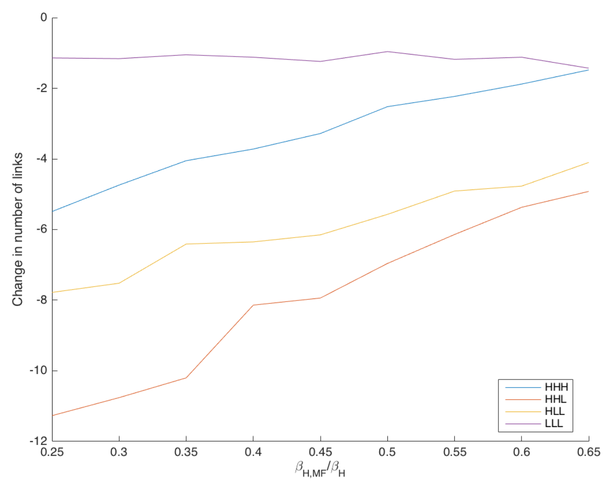
Figure D.1 presents the results. We plot the change in the number of links (and the change in the number of triangles) comparing MF networks to non-MF networks. We see that MF networks uniformly lead to a decline in every link and triangle type. Furthermore, the gap between the models declines the closer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

D.3.4. *Summary so far.* The simple models discussed so far with or without the idea of support all point to the same conclusion: that the number of LH and HH links should go down faster in MF villages than the number of LL links. There is however one additional factor, ignored so far, which might make the effect on the number of HH links in MF villages potentially ambiguous. This is the fact that microfinance itself promotes connections between group members, who will tend to be Hs (Feigenberg et al., 2013).

In addition, when we think of triads rather than pairs, our final model predicts that LLL triangles should be less likely to break apart in microfinance villages than LHH or LLH triangles. Because our data shows that LL links break the fastest (or at least faster than HH) and LLL triangles break the fastest (significantly more than mixed triads or HHH triads), none of these models can explain the data.



(A) Evolution of HH, LH, and LL



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

FIGURE D.1. Supported Links Model

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

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

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

In MF villages, observe that

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

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

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

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

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

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

We can see from the result that directed search will be inconsistent with the data namely because the effect on LL should be to increase, not decrease their presence, whereas the number of LH links will go down. Therefore, endogenous matching must generate spillovers in a another way.

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

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

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

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

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

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

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

The probability that a subgraph is maintained if it is already present is[46]

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

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

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

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

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

Online Appendix: Not for Publication

Appendix E. Hyderabad Network Elicitation

E.1. **Survey Questions.**

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

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

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

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

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

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

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

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

E.2. **ARD Algorithm.** We adapt the ARD algorithm from Breza et al. (2020)

We begin with a simple overview of the proposed method. Suppose that a researcher is interested in studying networks in a set of rural villages. A village network with n households is given by $\mathbf{g}$, which is a collection of links ij where $g_{ij} = 1$ if and only if households i and j are linked and $g_{ij} = 0$ otherwise. To fix ideas, suppose that the researcher wants to learn how some outcome variable W is related to a network statistic (or a vector of statistics) of interest $S(\mathbf{g})$. Or, perhaps the researcher is interested in how a treatment (such as exposure to microcredit) affects features of network structure, $S(\mathbf{g})$.

I. **Conduct ARD survey:** Sample a share ψ (e.g., 30 percent) of households. Have each enumerate a list of their network links.[47] Ask 5-8 ARD questions, such as

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

The ARD response for a household i is

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

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

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

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

Equipped with estimated fixed effects and latent locations, the probability of any network $\mathbf{g}$ being drawn is fully computed.

[47]Note that this gives a direct estimate of the respondent's degree. The method laid out in Section **??** does not require this and can also produce estimates for expected degree based on the ARD responses alone.

III. **Compute network statistics of interest:** Use the estimated probability model (using ζ, fixed effects ν_i and latent locations z_i) to compute $\mathrm{E}[S(\mathbf{g})|\mathbf{Y}]$. The code is freely available and discussed in Online Appendix B.[48]

IV. **Estimate economic parameter of interest:** E.g., run regressions such as

$$W_v = \alpha + \beta' \mathrm{E}[S(\mathbf{g}_v)|\mathbf{Y}_v] + \epsilon_v \text{ or } \mathrm{E}[S(\mathbf{g}_v)|\mathbf{Y}_v] = \alpha + \beta \mathrm{Treatment}_v + \epsilon_v,$$

[48]Note that here, the method produces estimates of the latent locations of each node, which may themselves be useful for some research questions.

Appendix F. Supplemental Borrowing Results

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

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

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

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

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

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

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

Panel A: H Nodes

	(1) MFI	(2) Informal (Non-Family)	(3) Friends	(4) SHG	(5) Moneylender	(6) Family
Microfinance × No. of Hs × Post	105.714	−142.299	−57.433	−6.290	−96.992	−82.186
	(17.362)	(88.130)	(23.670)	(28.833)	(80.457)	(71.299)
	[0.000]	[0.107]	[0.016]	[0.828]	[0.229]	[0.250]
MF × Post	2,465.432	−1,785.170	−1,413.222	236.424	−654.475	1,200.784
	(244.379)	(1,277.253)	(343.050)	(405.845)	(1,132.477)	(1,003.578)
	[0.000]	[0.163]	[0.000]	[0.561]	[0.564]	[0.232]
Depvar Mean	918.194	2295.64	879.686	2281.228	1426.571	2966.621
Observations	10,183	9,949	9,949	10,183	10,183	10,183

Panel B: L Nodes

	(1) MFI	(2) Informal (Non-Family)	(3) Friends	(4) SHG	(5) Moneylender	(6) Family
Microfinance × No. of Hs × Post	72.647	−135.154	−100.362	−87.567	95.696	−32.656
	(13.275)	(91.471)	(26.734)	(25.897)	(92.518)	(79.578)
	[0.000]	[0.140]	[0.000]	[0.001]	[0.301]	[0.682]
MF × Post	573.616	−3,003.094	−1,556.968	−759.694	385.459	−1,633.928
	(137.390)	(996.169)	(291.150)	(268.023)	(957.526)	(823.597)
	[0.000]	[0.003]	[0.000]	[0.005]	[0.688]	[0.048]
Depvar Mean	407.513	2501.673	935.019	1664.972	1566.634	2478.722
Observations	17,879	17,245	17,245	17,879	17,879	17,879

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

Appendix G. Results for Unsupported Links and Dependencies in Trangles

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

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

(where FIC is "Friends in Common")

Table G.1 presents the result.

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

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

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

G.2. **Dependencies in Maintenance of Triangles.** We assess the relevance of link interdependencies in our setting in the Karnataka panel data in non-microfinance villages. We consider triples of L nodes that are all connected in Wave 1 and ask whether those triangles are more likely to survive to Wave 2 than what we would expect under link independence. In Table G.1, we show that when LL links are unsupported (i.e., there is no third node with whom both members in the pair have a relationship), the probability of survival is 0.459. Thus, under independence, the survival probability for a full LLL trangle is $(0.459)^3 = 0.096$. However, we find that 24.2% of LLL triples observed in Wave 1 survive to Wave 2, which is substantially higher than what we would expect under independence. This suggests a potentially important role for local externalities and link dependencies.

Appendix H. Building Social Capital among MF Takers Cannot Explain Results

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

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

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

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

Table H.1. Link Evolution, Karnataka

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

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

Table H.2. Link Evolution, Karnataka

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

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

Appendix J. Misc. Calculation

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

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

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

and

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

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

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

and

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

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

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

Appendix K. Alternate Logistic Classification

K.1. Karnataka Exhibits.

Table K.1. Covariate balance, Karnataka

	Control			Treatment - control			
	Obs	Mean	SD	Coeff.	|Limit|	p-value	p-value 2
Distance to leaders > 1	7511	0.206	0.404	0.042	0.056	0.163	0.276
Harm mean distance to leaders	7511	0.474	0.079	-0.021	0.020	0.049	0.481
Eligible Female	7511	0.943	0.233	0.008	0.015	0.216	0.321
Leader	7511	0.154	0.361	-0.007	0.021	0.555	0.984
Share WC leader	7511	0.516	0.294	-0.060	0.078	0.112	0.461
WC harm mean dist	7511	0.495	0.186	-0.020	0.028	0.127	0.761
No electricity	7511	0.075	0.263	-0.018	0.022	0.134	0.239
No access to latrine	7511	0.748	0.434	-0.038	0.051	0.205	0.448
GMOBC	7511	0.709	0.454	-0.039	0.065	0.221	0.115
Num. rooms	7511	2.489	1.313	-0.001	0.140	0.973	0.502
Num. beds	7511	0.912	1.222	0.022	0.213	0.737	0.986
Thatched roof	7511	0.021	0.145	-0.002	0.014	0.678	0.815
RCC roof	7511	0.147	0.355	- 0.031	0.045	0.167	0.097
Household size	7511	5.014	2.205	0.297	0.231	0.013	0.049
Own rent	7511	1.303	0.988	-0.047	0.133	0.546	0.975
H	7511	0.752	0.432	0.005	0.042	0.800	0.300
Distance Blore	7511	61.114	17.458	-3.823	8.074	0.309	0.182
Distance Town	7511	5.647	3.595	1.203	1.862	0.190	0.142
All loans	7511	37861.564	129797.423	1351.740	11597.294	0.819	0.737
Network (friends and family) loans	7511	2735.470	25394.731	6.467	1716.401	0.994	0.788
SHG loans	7511	2543.994	6944.324	14.783	968.668	0.976	0.611
Bank loans	7511	19892.356	106358.225	3563.106	8808.589	0.428	0.316
Moneylender loans	7511	3638.339	20456.671	-164.660	1656.949	0.846	0.539

Notes: Classification of H type is based on logistic regression. This table reports Treatment-Control balance on predictors used to classify households as H (high MF propensity) vs. L (low MF propensity). Unit of observation: household. p-values of differences reflect standard errors clustered at the area level.
p-value 2: with village size control.

Table K.2. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance	−0.071	−0.081	−0.025	−0.023
	(0.023)	(0.023)	(0.010)	(0.008)
	[0.003]	[0.000]	[0.010]	[0.004]
Microfinance × LH	0.021	0.021	0.002	0.004
	(0.020)	(0.019)	(0.005)	(0.004)
	[0.315]	[0.269]	[0.631]	[0.293]
Microfinance × HH	0.022	0.022	0.005	0.008
	(0.025)	(0.021)	(0.007)	(0.005)
	[0.364]	[0.300]	[0.419]	[0.128]
LH	−0.068	−0.055	−0.012	−0.016
	(0.017)	(0.017)	(0.004)	(0.003)
	[0.000]	[0.002]	[0.006]	[0.000]
HH	−0.043	−0.025	−0.001	−0.012
	(0.019)	(0.019)	(0.006)	(0.005)
	[0.025]	[0.185]	[0.825]	[0.014]
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	0.523	0.523	0.0848	0.0848
MF + MF xLH= 0 p-val	0.017	0.007	0.001	0.002
MF + MF x HH = 0 p-val	0.006	0.002	0.005	0.016
MF + LHxMF = MF + HHxMF p-val	0.911	0.967	0.423	0.229

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

Table K.3. Link Evolution for Info and Financial Links, Karnataka

	(1) Financial Linked Post-MF	(2) Financial Linked Post-MF	(3) Info Linked Post-MF	(4) Info Linked Post-MF
Microfinance	−0.060	−0.013	−0.061	−0.016
	(0.027)	(0.007)	(0.026)	(0.007)
	[0.026]	[0.056]	[0.019]	[0.022]
Microfinance × LH	0.010	0.001	0.025	0.003
	(0.026)	(0.004)	(0.023)	(0.004)
	[0.690]	[0.800]	[0.270]	[0.499]
Microfinance × HH	0.013	0.002	0.017	0.003
	(0.028)	(0.005)	(0.027)	(0.005)
	[0.651]	[0.722]	[0.514]	[0.549]
LH	−0.067	−0.008	−0.069	−0.011
	(0.020)	(0.004)	(0.019)	(0.004)
	[0.001]	[0.042]	[0.000]	[0.003]
HH	−0.049	0.0001	−0.045	−0.003
	(0.021)	(0.004)	(0.022)	(0.004)
	[0.020]	[0.991]	[0.038]	[0.490]
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	0.415	0.0458	0.403	0.0534
MF + MF x LH= 0 p-val	0.024	0.003	0.055	0.002
MF + MF x HH= 0 p-val	0.012	0.008	0.009	0.003
MF + LHxMF = MF + HHxMF p-val	0.897	0.759	0.634	0.934

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

Table K.4. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance	−0.142	−0.132	−0.087	−0.084
	(0.041)	(0.030)	(0.041)	(0.030)
	[0.001]	[0.000]	[0.033]	[0.006]
Microfinance × LLH	0.072	0.054	0.027	0.020
	(0.032)	(0.026)	(0.030)	(0.024)
	[0.026]	[0.042]	[0.378]	[0.413]
Microfinance × LHH	0.105	0.078	0.036	0.023
	(0.041)	(0.032)	(0.042)	(0.032)
	[0.010]	[0.017]	[0.391]	[0.470]
Microfinance × HHH	0.107	0.068	0.044	0.023
	(0.042)	(0.032)	(0.044)	(0.032)
	[0.012]	[0.032]	[0.327]	[0.469]
LLH	−0.086	−0.064	−0.031	−0.015
	(0.021)	(0.022)	(0.016)	(0.016)
	[0.000]	[0.004]	[0.052]	[0.350]
LHH	−0.110	−0.071	−0.043	−0.010
	(0.026)	(0.028)	(0.024)	(0.024)
	[0.000]	[0.011]	[0.068]	[0.685]
HHH	−0.100	−0.052	−0.018	0.026
	(0.029)	(0.029)	(0.023)	(0.025)
	[0.001]	[0.077]	[0.444]	[0.292]
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	0.324	0.324	0.868	0.868
MF + MF x HHH= 0 p-val	0.147	0.004	0.019	0
MF + MF x LLH= 0 p-val	0.008	0.002	0.011	0.003
MF + MF x LHH= 0 p-val	0.142	0.034	0.015	0.001
MF + MF x HHH = MF + MF x LLH p-val	0.23	0.563	0.527	0.881
MF + MF xHHH= MF + MF x LHH p-val	0.936	0.595	0.661	0.993
MF + MF x LLH= MF + MF x LHHp-val	0.083	0.144	0.612	0.842

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

TABLE K.5. Borrowing patterns

	(1) MFI	(2) Informal (Non-Family)	(3) Friends	(4) SHG	(5) Moneylender	(6) Family
Microfinance × Post	266.259	−2,435.190	−779.130	−668.788	−719.414	666.310
	(135.746)	(841.210)	(238.425)	(244.851)	(790.856)	(686.540)
	[0.050]	[0.004]	[0.002]	[0.007]	[0.364]	[0.332]
Microfinance × Post × H	1,025.124	907.254	379.339	−294.704	647.986	−777.208
	(163.002)	(1,000.767)	(283.648)	(294.013)	(949.647)	(824.386)
	[0.000]	[0.365]	[0.182]	[0.317]	[0.496]	[0.346]
Microfinance × H	40.608	294.421	−72.549	−112.758	411.329	530.060
	(114.665)	(703.981)	(199.530)	(206.825)	(668.035)	(579.920)
	[0.724]	[0.676]	[0.717]	[0.586]	[0.539]	[0.361]
Post × H	169.662	−1,141.734	−685.428	528.206	−791.642	527.145
	(130.652)	(804.002)	(227.879)	(235.661)	(761.174)	(660.773)
	[0.195]	[0.156]	[0.003]	[0.026]	[0.299]	[0.426]
Observations	28,062	27,194	27,194	28,062	28,062	28,062
Depvar Mean	592.86	2426.289	914.773	1888.637	1515.799	2655.801
L, Non-MF Mean	171.561	3373.078	1360.849	1565.244	1809.136	3163.917
MF x Post x H + MF x Post =0 p-val	0	0.206	0.197	0.041	0.851	0.925

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

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

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

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

K.2. **Hyderabad Exhibits.**

TABLE K.7. Characteristics of H versus L, Hyderabad

Panel A: Hyderabad - Demographics and Amenities variables

	(1)	(2)	(3)	(4)	(5)
	GMOBC	Latrine	Num. Rooms	Thatched Roof	RCC Roof
H	0.027	0.090	0.055	0.006	−0.014
	(0.033)	(0.027)	(0.104)	(0.004)	(0.021)
	[0.417]	[0.001]	[0.596]	[0.129]	[0.504]
Depvar Mean	0.429	0.578	2.314	0.025	0.882
Observations	4,520	4,483	4,516	4,516	4,508

Panel B: Hyderabad - Network Variables

	(1)	(2)	(3)	(4)
	Expected Degree	Expected Links to L	Expected Links to H	Expected Centrality
H	0.758	0.004	0.749	0.017
	(0.221)	(0.249)	(0.196)	(0.004)
	[0.001]	[0.987]	[0.000]	[0.000]
Depvar Mean	5.813	10.06	-4.244	0.075
Observations	4,519	4,519	4,519	4,523

Notes: Classification of H type is based on logistic regression. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. The estimates reflect H versus L differences for the non-microfinance (control group) sample only.

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

	(1) Financial Links	(2) Financial Links	(3) Non Financial Links	(4) Non Financial Links
Microfinance	−0.155	−0.483	0.051	−0.114
	(0.151)	(0.139)	(0.155)	(0.145)
	[0.306]	[0.001]	[0.746]	[0.435]
Microfinance × H	−0.125	0.406	0.157	0.460
	(0.130)	(0.143)	(0.133)	(0.157)
	[0.338]	[0.006]	[0.238]	[0.005]
H	0.144	−0.489	−0.086	−0.451
	(0.125)	(0.147)	(0.126)	(0.159)
	[0.255]	[0.002]	[0.500]	[0.006]
Observations	4,429	4,429	4,429	4,429
Double-Post LASSO	No	Yes	No	Yes
Depvar Mean	4.24	4.24	2.87	2.87
MF + MF × H = 0 p-val	0.16	0.683	0.314	0.082

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

Table K.9. Link Evolution, Hyderabad

	(1) Prob. Linked	(2) Prob. Linked
Microfinance	−0.001	−0.001
	(0.002)	(0.003)
	[0.583]	[0.716]
Microfinance x HH	−0.017	−0.016
	(0.007)	(0.007)
	[0.017]	[0.015]
Microfinance x LH	−0.006	−0.006
	(0.003)	(0.002)
	[0.019]	[0.011]
HH	0.022	0.023
	(0.006)	(0.006)
	[0.000]	[0.000]
LH	0.007	0.007
	(0.002)	(0.002)
	[0.002]	[0.000]
Observations	141,990	141,990
Controls	No	Yes
Depvar Mean	0.0255	0.0255
LL, Non MF Mean	0.0246	0.0246
MF + MF x $HH = 0$ p-val	0.016	0.019
MF + MF x $LH = 0$ p-val	0.04	0.049
MF + MF x HH = MF + MF x LH p-val	0.027	0.031

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

Table K.10. Triples Evolution, Hyderabad

All variables x 1000	(1) Full Triangle Linked	(2) Full Triangle Linked
Microfinance	0.00	0.01
	(0.02)	(0.02)
	[0.853]	[0.817]
Microfinance x LLH	-0.1	-0.1
	(0.01)	(0.01)
	[0.000]	[0.000]
Microfinance x LHH	-0.1	-0.1
	(0.05)	(0.04)
	[0.006]	[0.006]
Microfinance x HHH	-0.3	-0.3
	(0.1)	(0.1)
	[0.011]	[0.013]
LLH	0.03	0.03
	(0.01)	(0.01)
	[0.004]	[0.003]
LHH	0.1	0.1
	(0.05)	(0.04)
	[0.005]	[0.005]
HHH	0.3	0.3
	(0.1)	(0.1)
	[0.009]	[0.010]
Observations	3,341,002	3,341,002
Controls	No	Yes
Depvar Mean	6.82e-02	6.82e-02
LLL, Non-MF Mean	5.2e-02	5.2e-02
MF + MF x $HHH = 0$ p-val	0.013	0.016
MF + MF x $LLH = 0$ p-val	0.028	0.036
MF + MF x $LHH = 0$ p-val	0.016	0.017
MF + MF x HHH = MF + MF x LLH p-val	0.024	0.03
MF + MF x HHH = MF + MF x LHH p-val	0.025	0.03
MF + MF x LLH = MF + MF x LHH p-val	0.049	0.066

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

Table K.11. Borrowing patterns, Hyderabad

	(1) MFI	(2) Informal (Non-Family)	(3) Friends	(4) SHG	(5) Moneylender	(6) Family
Microfinance	1,385.829	−6,320.583	−1,321.914	−2,107.269	−2,913.419	−215.076
	(517.256)	(2,870.182)	(1,083.853)	(1,045.468)	(2,029.758)	(922.000)
	[0.008]	[0.028]	[0.223]	[0.044]	[0.152]	[0.816]
Microfinance × H	1,581.823	471.937	1,227.309	−363.506	−9.592	938.285
	(504.193)	(3,158.391)	(1,258.093)	(1,385.831)	(2,186.468)	(928.055)
	[0.002]	[0.882]	[0.330]	[0.794]	[0.997]	[0.313]
H	−1,512.372	−1,972.374	−1,832.799	57.352	−556.518	−949.171
	(501.665)	(3,119.204)	(1,289.006)	(1,352.461)	(2,213.463)	(892.681)
	[0.003]	[0.528]	[0.156]	[0.967]	[0.802]	[0.288]
Depvar Mean	3138.51	35298.64	8331.35	7162.35	19331.47	2698.52
L, Non MF Mean	2095.07	35626.91	8200.6	7142.77	19884.52	2683.8
MF + MF × H = 0 p-val	0	0.166	0.95	0.165	0.287	0.567
Observations	6,811	6,863	6,863	6,863	6,863	6,863

Notes: Classification of H type is based on logistic regression. These tables present the effect of microfinance access on the loan amounts borrowed from various sources. Outcomes are winsorized to the 1% level. Here all specifications include demographic household and village controls (the same ones used in random forest classification of H vs L) subject to double-post LASSO. Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets.

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

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

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

Table K.13. Risk sharing, Hyderabad

	(1) Expenditures: Total	(2) Expenditures: Non-Food
Household Income per capita	0.110	0.055
	(0.032)	(0.025)
	[0.001]	[0.026]
Microfinance × Income	0.133	0.109
	(0.075)	(0.066)
	[0.077]	[0.098]
Household Income per capita $\times H$	0.042	0.016
	(0.062)	(0.045)
	[0.491]	[0.717]
Microfinance × Income $\times H$	-0.113	-0.066
	(0.109)	(0.088)
	[0.304]	[0.450]
Observations	10452	10361
Depvar Mean	1192.937	1192.937
L, Non MF Depvar Mean	1184.050	1184.050
Income Mean	1436.279	1439.993
L, Non MF Income Mean	1433.961	1436.304
Test: MF × Income + MF × Income $\times H$	0.806	0.506

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

Appendix L. Hyderabad Robustness

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

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

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

Appendix M. Karnataka Robustness

Table M.1. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF	(5) Linked Post-MF	(6) Linked Post-MF
Microfinance	−0.058	−0.038	−0.028	−0.023	−0.005	−0.005
	(0.018)	(0.018)	(0.018)	(0.008)	(0.006)	(0.006)
	[0.002]	[0.039]	[0.120]	[0.006]	[0.362]	[0.368]
Microfinance × LH	0.009	0.007	−0.004	0.007	0.007	0.005
	(0.015)	(0.015)	(0.017)	(0.004)	(0.003)	(0.003)
	[0.573]	[0.635]	[0.832]	[0.120]	[0.043]	[0.088]
Microfinance × HH	0.039	0.038	0.015	0.009	0.009	0.011
	(0.022)	(0.020)	(0.020)	(0.007)	(0.005)	(0.005)
	[0.086]	[0.059]	[0.467]	[0.206]	[0.073]	[0.044]
LH	−0.025	−0.030	−0.127	−0.002	−0.006	0.050
	(0.012)	(0.012)	(0.297)	(0.004)	(0.003)	(0.080)
	[0.036]	[0.018]	[0.670]	[0.566]	[0.020]	[0.530]
HH	0.008	−0.001	−0.147	0.021	0.012	0.138
	(0.017)	(0.015)	(0.472)	(0.006)	(0.004)	(0.126)
	[0.622]	[0.942]	[0.755]	[0.001]	[0.001]	[0.273]
Observations	57,376	57,376	57,376	846,561	846,561	846,561
Linked Pre-MF	Yes	Yes	Yes	No	No	No
Village size control		✓	✓		✓	✓
Village size interacted with LH,HH control			✓			✓
Depvar Mean	0.441	0.441	0.441	0.0636	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.482	0.0753	0.0753	0.0753
MF + MF xLH= 0 p-val	0.014	0.179	0.194	0.015	0.852	0.972
MF + MF xHH= 0 p-val	0.361	0.994	0.542	0.101	0.615	0.4
MF + LHxMF = MF +HHxMF p-val	0.137	0.1	0.409	0.641	0.612	0.211

Notes: Standard errors (clustered at the village level) are reported in parentheses. p-values are reported in brackets. Village controls are village size, village size^2 and village size^3. We also include interactions with LH and HH.

Table M.2. Link Evolution, Karnataka

	(1) Linked Post-MF	(2) Linked Post-MF	(3) Linked Post-MF	(4) Linked Post-MF
Microfinance	−0.058	−0.160	−0.023	−0.003
	(0.018)	(0.121)	(0.008)	(0.023)
	[0.002]	[0.187]	[0.006]	[0.885]
Microfinance × LH	0.009	−0.004	0.007	0.008
	(0.015)	(0.015)	(0.004)	(0.005)
	[0.573]	[0.796]	[0.120]	[0.077]
Microfinance × HH	0.039	0.015	0.009	0.014
	(0.022)	(0.024)	(0.007)	(0.007)
	[0.086]	[0.512]	[0.206]	[0.039]
LH	−0.025	−0.002	−0.002	−0.008
	(0.012)	(0.012)	(0.004)	(0.004)
	[0.036]	[0.864]	[0.566]	[0.068]
HH	0.008	0.046	0.021	0.006
	(0.017)	(0.020)	(0.006)	(0.007)
	[0.622]	[0.023]	[0.001]	[0.358]
Observations	57,376	57,376	846,561	846,561
Linked Pre-MF	Yes	Yes	No	No
Controls		✓		✓
Depvar Mean	0.441	0.441	0.0636	0.0636
LL, Non-MF Mean	0.482	0.482	0.0753	0.0753
MF + MF xLH= 0 p-val	0.014	0.18	0.015	0.843
MF + MF x HH = 0 p-val	0.361	0.244	0.101	0.652
MF +LHxMF = MF + HHxMF p-val	0.137	0.344	0.641	0.195

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

Table M.3. Triples Evolution, Karnataka

	(1) Full triangle linked Post-MF	(2) Full triangle linked Post-MF	(3) Any link in triangle survived Post-MF	(4) Any link in triangle survived Post-MF
Microfinance	−0.078	−0.166	−0.085	−0.077
	(0.029)	(0.150)	(0.023)	(0.128)
	[0.008]	[0.269]	[0.000]	[0.548]
Microfinance × LLH	0.026	−0.002	0.043	0.023
	(0.021)	(0.022)	(0.018)	(0.016)
	[0.228]	[0.946]	[0.015]	[0.149]
Microfinance × LHH	0.054	−0.004	0.057	0.021
	(0.030)	(0.029)	(0.025)	(0.019)
	[0.072]	[0.896]	[0.022]	[0.282]
Microfinance × HHH	0.093	0.002	0.087	0.037
	(0.042)	(0.040)	(0.031)	(0.028)
	[0.028]	[0.953]	[0.006]	[0.194]
LLH	−0.024	0.007	−0.037	−0.018
	(0.018)	(0.020)	(0.014)	(0.014)
	[0.180]	[0.718]	[0.009]	[0.203]
LHH	−0.037	0.030	−0.032	0.006
	(0.025)	(0.025)	(0.017)	(0.014)
	[0.133]	[0.242]	[0.053]	[0.696]
HHH	−0.025	0.076	−0.012	0.039
	(0.033)	(0.032)	(0.022)	(0.021)
	[0.454]	[0.020]	[0.593]	[0.068]
Observations	53,233	53,233	53,233	53,233
Linked Pre-MF	Yes	Yes	Yes	Yes
Controls		✓		✓
Depvar Mean	0.197	0.197	0.808	0.808
LLL, Non-MF Mean	0.252	0.252	0.864	0.864
MF + MF x HHH = 0 p-val	0.698	0.303	0.935	0.762
MF + MF x LLH= 0 p-val	0.023	0.273	0.022	0.686
MF + MF x LHH= 0 p-val	0.262	0.28	0.141	0.666
MF + MF x HHH = MF + MF x LLH p-val	0.076	0.909	0.093	0.604
MF + MF x HHH = MF + MF x LHH p-val	0.212	0.83	0.075	0.369
MF + MF x LLH= MF + MF x LHH p-val	0.122	0.894	0.409	0.868

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

Appendix N. Model Simulation

We provide a simulation with the following parameters to demonstrate that LL links can drop more than HL or HH links. The parameters are as follows:

- $\lambda_H = 0.25$
- base benefits and costs of socializing:
  - $c = 50$
  - $u = 0.05$
- values for links chosen
  - $v_{\theta\theta'} = 2$ for $\theta\theta' \neq LH$ and $v_{LH} = 5$
  - common for simplicity, with increased value to link with Hs for Ls
- impact of microfinance: only reduces value of L to H and nothing more:
  - $v_{HL}^{mf} = 0.99 \cdot v_{HL}$
  - $v_{\theta\theta'}^{mf} = v_{\theta\theta'}$ if $\theta\theta' \in \{HH, LH, LL\}$

With these parameters we can estimate equilibrium efforts by $e = (I - E)^{-1}$ as well as the vector of degrees $(d_{HH}, d_{HL}, d_{LH}, d_{LL})$ and corresponding densities.

Table N.1. Parameter Estimates

	No Microfinance	Microfinance	Difference
d_{HH}	3.7780	3.1515	-0.6266
d_{HL}	14.7859	12.3862	-2.3997
d_{LH}	4.9286	4.1287	-0.7999
d_{LL}	18.4213	15.5322	-2.8891
density_{HH}	0.0756	0.0630	-0.0125
density_{HL}	0.0986	0.0826	-0.0160
density_{LH}	0.0986	0.0826	-0.0160
density_{LL}	0.1228	0.1035	-0.0193
e_H	0.2813	0.2569	-0.0244
e_L	0.3586	0.3293	-0.0293

In the simulation, the average degrees are comparable to those in the data. Microcredit introduces roughly a 1-2pp decline in the probability of linking on a base of density roughly 10%. We observe that efforts of both H and L decline, but particularly more by Ls in absolute terms. The maximal decline in links comes from LLs both in terms of absolute link count and also density.

Appendix O. Karnataka Links to New Households

Table O.1. Link Evolution, Karnataka

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

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