AN EXPERIMENTAL TEST OF THE ASSOCIATION BETWEEN NETWORK CENTRALITY AND CROSS-VILLAGE RISK-SHARING LINKS

ERICA FIELD[§] AND ROHINI PANDE[¶]

ABSTRACT. We test a prediction from a recent paper by Ambrus and Elliott (2018), according to which less volatile incomes increases the association between within community centrality of a household, defined as Myerson centrality, and the probability of keeping financial connections with households outside the village. We use data from a unique field experiment in 185 Indian villages in which a randomly chosen half of the villages got access to formal banking services. We find empirical support for the prediction, as the relationship between Myerson centrality and having outside links is significantly more positive in villages that got access to formal banking.

Date: August 14, 2018.
We use algorithms developed in conjunction with Arun Chandrasekhar to facilitate our empirical strategy, and are very grateful for his intellectual contribution. These algorithms are described in Appendix B.
[§]Department of Economics, Duke University; email: emf23@duke.edu.
[¶]Harvard Kennedy School of Government; email: rohini_pande@ksg.harvard.edu.

1. Introduction

In a recent paper Ambrus and Elliott (2018) propose a strategic model of network formation in a context of informal risk-sharing. The model features distinct groups of households, with the central assumption that it is costlier to maintain connections across groups than within, but also that it is potentially more beneficial, because incomes are less positively correlated across groups than within. The model predicts that if maintaining across group links is relatively costly, households typically underinvest in across-group connections relative to what would be socially optimal, as those maintaining these connections do not take into account the positive externality the connection generates for their group members. It is also shown that the households that have the highest private incentives to invest into outside connections are the more central ones, according to a new measure of network centrality they label Myerson centrality. Lastly, the model predicts that when household face less volatile incomes and therefore the value of informal risk-sharing connections decreases, the association between Myerson centrality and maintaining across group connections should be higher since less central households do not find it worthwhile anymore to maintain such connections.

We test the latter theoretical prediction using unique network data from 185 villages in Tamil Nadu, India. The data were collected in the context of a large field experiment in which a randomly chosen half of the villages gained access to local banking services, providing exogenous variation in access to formal loans. Both in the treated and non-treated villages, near complete network data were collected on within-village actual and potential financial transactions, as well as on financial links to households outside the village. Using these data, we test the theoretical prediction that the association between Myerson centrality within the village network and having financial links outside the village becomes more positive when villagers have access to formal banking (and therefore the value of informal financial links is smaller). This prediction bears out in the data. In particular, the relationship between Myerson centrality and outside links is significantly more positive in villages that were randomly chosen to receive formal banking services. This prediction can confidently be interpreted as a causal influence of a reduction in the value of outside links because it is based on truly exogenous variation in formal credit access, which gives rise to a significant reduction in the number of network links, consistent with a decline in the relative importance of informal risk-sharing.

Our dataset is particularly well-suited for our analysis as it (i) involves numerous independent villages (essential for inference, though most network-based studies have just one or a handful of villages), (ii) includes complete network data across both financial and social connections for almost all households in every village, and (iii) captures both within village contacts and outside-village contacts, which are rarely contained in datasets of this kind. However, the major advantage of these data for testing the empirical predictions of our model is that they were collected in the context of a large field experiment in village banking, in which a randomly chosen half of villages gained access to local banking services

1-2 years prior to data collection. This gives us indisputably exogenous variation in access to formal financial services with which to test one of the model's key predictions: that the correlation between Myerson Centrality and number of costly links is more positive when the value of outside risk-sharing links is lower, all else equal. The setting is unique in the sense that reliance on informal networks varies randomly with the introduction of banking services, which allows us to study how truly exogenous variation in the value of network links across villages influences network composition.

Our work contributes to a recent string of papers examining how social networks respond to the introduction of various financial instruments: see for example Feigenberg et al. (2013), Banerjee et al. (2014a, 2014b) and Binzel et al. (2017). More broadly the paper is related to the empirical literature on informal risk-sharing arrangements.[1]

2. Setting and Data

The data we employ were collected from 2014 to 2016 in conjunction with a large-scale impact evaluation of access to formal financial services in rural Tamil Nadu, India (Binzel et al., 2017). The implementing partner was a large financial institution (henceforth, LFI) that offers group-based and individual loans to both men and women through local village branches with the explicit goal of reaching individuals in financially marginalized (previously unbanked) rural communities. Beginning in 2008, LFI expanded bank infrastructure across villages from the districts of Thanjavur, Thiruvarur and Pudukkotai (Tamil Nadu). Prior to this rolling-out, 102 potential branch service areas (henceforth, SAs) were identified by LFI as potential expansion areas. The average SA spans 10 villages within a radius of 4-5 km from the branch and covers a population of roughly 10,000 people. Once all feasible branch locations in the district had been designated, SAs were matched into pairs using a minimum distance matching algorithm, and 51 bank branches were randomly assigned to one SA in each pair.

Bank operations began soon after treatment assignment. By the onset of network data collection efforts, bank penetration had reached a average level of 41% in treatment SAs. An early evaluation of this expansion of financial activity conducted in 2013 shows that households living in treatment villages were 32% more likely to have borrowed in the previous year compared with households in control villages.

Beginning in 2014, a full social network mapping survey was administered in a randomly chosen subset of 204 villages from the 102 service areas (2 villages per service area).[2] In these

---

[1]For an incomplete list of papers see Ellsworth (1988), Rosenzweig (1988), Deaton (1992), Paxson (1993), Udry (1994), Townsend (1994), Grimard (1997), Fafchamps and Lund (2003), Schulhofer-Wohl (2011) and Mazzocco and Saini (2012).

[2]At baseline, i.e. before branch openings in treatment service areas, two villages per service area were selected as follows. First, the sample was limited to villages with 40-250 households, excluding the designated branch location. For each pair, one village was randomly selected and then matched with the village in the corresponding treatment or control service area that had as close to the same distance from its respective branch as the first picked village had from its branch. For control area villages, the planned branch location was used as benchmark.

villages, all households were asked to name all social and financial contacts both within and outside the village, enabling us to map the full network of social and financial connections within each sampled village.[3] Households were surveyed 18 to 24 months after the opening of the branch.[4] The network survey was administered to both the head and spouse of each household, when available.[5] In the survey, the head of the household and spouse were asked to identify all individuals within the village with whom they: (i) spend leisure time; (ii) could borrow in case of emergency; and (iii) could borrow to finance a business investment.[6] In addition, respondents were asked to list all individuals living *outside* of the village from whom they could borrow in case of emergency. In addition to naming each link both inside and outside the village, respondents were asked their relationship to the link (friend, family, employer, moneylender), the actual amount borrowed from each link, the amount they could borrow from each link in case of emergency, the amount they could borrow from each link to finance a business investment, and the number of contact days with each link (out of past 7). For outside links, respondents were also asked the distance to each link (for our purposes, whether the link lives within walking distance). Information on outside links was only collected in 189 villages, and village-level controls are missing from 4 villages. As a consequence, this analysis is limited to the 185 villages with complete data.[7] We also exclude the 38 households that moved into the study area between baseline and endline.

Although financial and leisure ties are elicited separately for both the head and spouse, in order to analyze household-level networks we aggregate observations within the household in the following manner. First, we only consider "OR" networks - that is, those containing either a social or a financial tie.[8] Second, we aggregate the two layers of edges between households using the following rule: If two people from household A report two persons in household B, there is a *unique* directed edge from A to B. Thus, it will be equivalent to the case where only one person from household A reports a person from household B.

[3]Links named by respondents were immediately matched to names within a database of village members collected at baseline. Information on outside contacts cannot be mapped since household living in villages that are not included in our sample cannot be identified by name and location.

[4]In 85 villages, an additional round of network data was also collected at Baseline (prior to the opening of the bank branch) in addition to Endline. Because less than half of villages have panel network data, baseline data are excluded from the current analysis.

[5]For 23% of the households, only the head of the household has been interviewed. For 13% only the spouse has been interviewed.

[6]In all questions, households could list up to 15 individuals, which results in very little censoring of networks. The maximum number of links was reached in only very few cases (less than 0.01% of cases).

[7]Inclusion in the sub-sample is balanced across treatment and control.

[8]In particular, our analysis focuses on two types of networks: the *financial graph* L^f and the *social graph* L^s. The financial graphs represent risk-sharing connections, and the social graph represents friendships and ties used to socialize (see survey questionnaire in Section A.1 of the Supplementary Appendix), which are not mutually exclusive. Our empirical test utilizes both types of links and considers L^{all} the network of either risk-sharing or friendship connections. We favor "OR" networks because of the high degree of overlap between social and financial networks, and because of the concern that network links are self-censored due to imperfect recall or insufficient recall effort. The results are almost identical when MC is computed on the financial or social networks alone.

Second, to aggregate characteristics of the interaction between two households, we consider an aggregation rule that avoids double-counting. The continuous value characterizing a link between household A and household B is the maximum of all the values that the head and the spouse of household A have reported for *anyone* belonging to household B.[9]

These aggregation rules are only applicable for inside village contacts in which we know whether two declared contacts belong to the same household. For outside contacts, we consider all the outside contacts listed by the household, excluding only contacts that are classified by respondents as money-lenders.

3. Results on Consumption Variance and Myerson Centrality

Tables 2-3 provide summary statistics of the sampled villages in order to verify that the sample is balanced across treatment and control arms. The average number of households per village is 112, the average node degree is 4.56, and the density is 0.04. Tables 4-6 show the treatment effects of village-level banking services on within-village and outside-village links, which are discussed extensively in Binzel et al. (2017). As predicted, the introduction of banking services generates an exogenous reduction in households' within-village network links and their reliance on informal transfers, as measured by the difference in real and potential borrowing levels between treatment and control villages (Table 4).

In Tables 5 and 6 we observe that banking does not reduce the number of outside links, but does lead to less informal borrowing from outside sources, and to changes in the composition of outside links. In particular, in banked villages, outside links are younger, closer in distance, and more social than outside links in control villages. All of these suggest a shift away form more financially valuable outside links.

Overall, these results demonstrate that the randomized introduction of formal banking services reduced the value of informal links both within and outside the village, as would be expected. This allows us to rigorously test the more nuanced predictions on the relationship between bridging links and Myerson Centrality. Our main test of the theoretical model, presented in Table 1, utlizes the following specification:

$$\#OutContact_{ji} = \alpha_0 + \alpha_1 T_i + \alpha_2 MC_{ji} + \alpha_3 MC_{ji} \times T_i + \alpha_4 X_{ij} + \gamma_{s_i} + \epsilon_{it} \tag{1}$$

Where $\#OutContact_{ij}$ is a binary indicator of whether household j in village i has any links outside the village, T_i is the treatment indicator equals to 1 if village i was in a service area that was randomly given access to the LFI's services, MC_{ji} is the Myerson Centrality of individual j computed on the ALL network, X_{ij} is a set of control variables either at the household or the village level, γ_{s_i} is a pair fixed effect that accounts for the experimental stratification, and ϵ_{it} is an error term. Since the treatment is assigned at the service area level (encompassing several villages), this error term is clustered at the service area level.

[9]For instance, if the head of household A report that she can borrow Rs. 150 from the spouse of household B and the spouse of household A reports that he can borrow Rs. 100 from the head of the household B, we will consider that household A can borrow Rs. 150 from household B.

In Table 1 we see clear evidence that an exogenous reduction in the value of outside links brought about by the introduction of formal banking services leads to a significantly more positive relationship between Myerson Centrality and link formation. The negative and significant coefficient estimate on the treatment indicator implies that banking services encouraged an overall reduction in outside links for all individuals in the village. Meanwhile, the positive and significant coefficient estimate on the interaction between MC and treatment implies that the impact of access to formal banking on outside links was less extreme among more central individuals. That is, although outside links are less common in the new regime in which those links are less financially valuable, Myerson central individuals are significantly more likely to retain outside links compared to less central individuals when their value declines.[10] Appendix A.3 also shows that villagers' incomes are positively correlated with their Myerson centralities, as predicted by the theory.

[10]The negative coefficient on Myerson centrality is not predicted by the theory, but it is also not inconsistent with it: when the benefits of across group links are large then all agents, including those who are not central in their own group's network, have incentives to establish outside links, and our model does not give sharp predictions on the structure of the network. Furthermore, there might be individuals who have more recently moved to the village and are both more likely to have "free" outside links and to be less central within the village community. Unfortunately, we do not have information on how long a household has resided in the village to test this hypothesis.

Table 1. Treatment effect on whether household has outside contacts

	Has any outside contact (1)	Has any outside contact (excluding contacts within walking distance) (2)	Has any outside contact (3)	Has any outside contact (excluding contacts within walking distance) (4)	Has any outside contact (5)	Has any outside contact (excluding contacts within walking distance) (6)
Treatment	-0.0310*	-0.0288	-0.0274*	-0.0255	-0.0512*	-0.0614**
	(0.0165)	(0.0191)	(0.0157)	(0.0176)	(0.0285)	(0.0270)
MC	-0.000201***	-0.000156***	-0.000108***	-0.0000880**	-0.000415	-0.000920*
	(0.0000469)	(0.0000528)	(0.0000306)	(0.0000339)	(0.000419)	(0.000501)
MC × Treatment	0.000113**	0.0000830	0.0000788***	0.0000558*	0.000433*	0.000463**
	(0.0000449)	(0.0000531)	(0.0000294)	(0.0000332)	(0.000221)	(0.000197)
Nr. Observations	18,648	18,648	18,648	18,648	18,648	18,648
Nr. villages	185	185	185	185	185	185
R2	0.164	0.143	0.163	0.142	0.162	0.142
MC Calculation	Method 1, Undirected	Method 1, Undirected	Method 2, Undirected	Method 2, Undirected	Method 3, Undirected	Method 3, Undirected
Mean Control	0.60	0.52	0.60	0.52	0.60	0.52

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses
Control variables include, at the household level, a dummy of whether the only respondent of the household was the head of the household, a dummy of whether the only respondent of the household was the head of the household's spouse, the average age of the household's respondents, a dummy of whether the household is involved in agriculture. Control variables at the village level are: the number of households in the village, the distance to the bank branch, and the proportion of people belonging to the same caste in the village. Control variables take a value of zero when missing values, and regressions include an indicator of missing data corresponding to each control. Control *HH* refers to a regression including only the first two dummies of the household controls.

References

Algaba, E., JM Bilbao, J. Fernández, N. Jiménez and J. López (2007): "Algorithms for computing the Myerson value by dividends," Discrete Mathematics Research Progress, 1-13.

Ambrus, A. and M. Elliott (2018): "Investments in Social Ties, Risk Sharing and Inequality," *mimeo Duke University and University of Cambridge.*

Banerjee, A., A. Chandrasekhar, E. Duflo and M. Jackson (2013a): "The diffusion of microfinance," Science, 341, No. 6144 1236498

Banerjee, A., A. Chandrasekhar, E. Duflo and M. Jackson (2014a): "Exposure to Formal Loans: Changes in Social Networks," *mimeo MIT and Stanford University.*

Banerjee, A., E. Breza, E. Duflo and C. Kinnan (2014b): "Does Microfinance Foster Business Growth? The Importance of Entrepreneurial Heterogeneity," *mimeo MIT.*

Binzel, C., E. Field and R. Pande (2017): "Does the Arrival of a Formal Financial Institution Alter Informal Sharing Arrangements? Experimental Evidence from Village India," *mimeo Harvard University.*

A. Chandrasekhar and R. Lewis (2014): "Econometrics of sampled networks," mimeo Stanford University.

Deaton, A. (1992): "Household Savings in LDCs: Credit Markets, Insurance and Welfare," *The Scandinavian Journal of Economics*, 94, 253-273.

Ellsworth, L. (1988): "Mutual Insurance and Non-market Transactions among Farmers in Burkina Faso," *PhD. thesis, University of Wisconsin, Madison.*

Feigenberg, B., E. Field and R. Pande (2013): "The Economic Returns to Social Interaction: Experimental Evidence from Micro Finance," *Review of Economic Studies*, 80, 1459-1483.

Fafchamps, M. and S. Lund (2003): "Risk-sharing Networks in Rural Philippines," *Journal of Development Economics*, 83, 326-350.

Grimard, F. (1997): "Household Consumption Smoothing through Ethnic Ties: Evidence from Cote d'Ivoire," *Journal of Development Economics*, 53, 391-422.

Mazzocco, M. and S. Saini (2012): "Testing Efficient Risk Sharing with Heterogeneous Risk Preferences," *American Economic Review*, 102, 428-468.

Paxson, C. H. (1993): "Consumption and Income Seasonality in Thailand," *Journal of Political Economy*, 101, 39-72.

Rosenzweig, M. R. (1988): "Risk, Implicit Contracts and the Family in Rural Areas of Low-income Countries," *The Economic Journal*, 98, 1148-1170.

Schulhofer-Wohl, S. (2011): "Heterogeneity and Tests of Risk Sharing," *The Journal of Political Economy*, 119, 925-958.

Townsend, R. M. (1994): "Risk and Insurance in Rural 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.

Appendix A

A.1. **Survey questions and definition of networks.** As part of the social and informal risk-sharing network mapping, we asked about three types of ties: inside their village leisure contact, inside their village borrowing contact and outside their village contact.

Inside leisure contacts were elicited by asking respondents to "think of the people, within their gramum with whom *they* spend the most leisure time (non-household members above the age of 18). These are people with whom *they* may have spent time for your relaxation, during breaks at work, discussing in person or on the phone, at festivals, drinking tea, or whenever *they* have made time for yourself (free time)".

Additional borrowing contacts within the village were elicited by asking respondents to list "additional people (outside of their household) who *they* could borrow from in case of emergency, other than those people *they* have already listed".

Finally, contacts living outside the village were elicited by asking respondents to "list people outside the gramum from whom *they* could borrow in the case of an emergency".

For each contact, we asked the respondent to specify:

(1) The number of days in the last 7 when you have met or spent time with this person face-to-face or spoken to this person on the phone.
(2) The total amount of money actually borrowed from this person in the last 12 months.
(3) The maximum amount of money that this person would have been willing to lend you over the last 12 months.
(4) The amount of money the respondent could borrow from this person if she was going to start a business or expand an existing business, over the past 12 months.

For outside contacts we also ask where the contact lives.
Based on this set of question, we define two types of inside ties:

- Leisure contact are any household listed as an *inside leisure* contact.
- Borrowing contact are any household listed as an *inside leisure* contact or as an *inside borrowing* contact such as one of the answer to questions 2, 3 or 4 is strictly positive, i.e. it is a contact the respondent can either borrow money from, has already borrowed from or could borrow from in case she wants to start a business.

Finally, from these two types of links we draw the final graphs used in the analysis:

- *Bidirected financial network:* there is a link between household A and B iif A declares B and B declares A as a borrowing contact.
- *Bidirected social network:* there is a link between household A and B iif A declares B and B declares A as a leisure contact.

- *Bidirected all network:* there is a link between household A and B iif A declares B as either a leisure or a borrowing contact and B declares A as either a leisure or a borrowing contact.

A.2. **Tables.**

Table 2. Balance statistics: Village characteristics at Endline - Villages with more than 40 households

	Observations $N_C + N_T$ [1]	**Control Mean [SD]** [2]	**Treatment Mean diff. (SE)** [3]
Village characteristics			
Number of households (Census)	N= 185	119.52	-8.771
	93C + 92T	[55.54]	(6.688)
Number of heads and spouses (Census)	N= 185	214.77	-16.793
	93C + 92T	[99.04]	(11.788)
Number of surveyed households (SNM)	N= 185	110.71	-9.974
	93C + 92T	[53.47]	(6.479)
Number of surveyed heads and spouses (SNM)	N= 185	175.51	-18.774*
	93C + 92T	[87.80]	(9.829)
Pct. of surveyed households	N= 185	0.92	-0.006
	93C + 92T	[0.09]	(0.010)
Pct. of surveyed heads and spouses	N= 185	0.81	-0.011
	93C + 92T	[0.10]	(0.009)
Population estimate (Indian Census, 2001)	N= 185	488.83	-34.688
	93C + 92T	[213.33]	(25.550)
Distance to the bank branch, *kms*	N= 185	2.31	-0.051
	93C + 92T	[1.32]	(0.114)

Note : The sample is restricted to villages with outside contact information. . ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively. Column (1) reports the total number of observation and its decomposition by groups. Column (2) reports the average outcome (standard deviation) for the control group. Column (3) reports the regression coefficent associated to the treatment dummy when controlling for pair fixed effects and with error terms cluster at the Service Area level.

Table 3. Descriptive Statistics - Undirected networks

	N	Mean (SD)	Min-Max
	[1]	[2]	[3]
Average Myerson Centrality - Incoming link deletion algorithm (Method 1)			
Leisure network	185	169.49 (142.91)	0.77 - 569.05
Financial network	185	188.02 (167.11)	3.67 - 768.19
All network	185	227.90 (184.56)	6.37 - 805.90
Std Dev Myerson Centrality -Incoming link deletion algorithm (Method 1)			
Leisure network	185	35.71 (27.11)	0.84 - 212.86
Financial network	185	42.82 (27.98)	2.92 - 149.75
All network	185	44.35 (26.67)	3.98 - 140.46
Average Myerson Centrality - Outgoing link detection algorithm (Method 2)			
Leisure network	185	236.11 (207.57)	0.64 - 801.46
Financial network	185	258.25 (243.28)	2.27 - 1189.14
All network	185	315.79 (269.74)	4.77 - 1237.11
Std Dev Myerson Centrality - Outgoing link detection algorithm (Method 2)			
Leisure network	185	65.58 (63.47)	1.01 - 546.02
Financial network	185	71.86 (59.44)	1.72 - 329.28
All network	185	77.95 (59.94)	3.09 - 321.49
Average Myerson Centrality - Link detection algorithm (Method 3)			
Leisure network, bidirected	185	56.49 (39.36)	13.67 - 215.75
Financial network	185	53.78 (31.93)	13.58 - 181.04
All network	185	50.22 (29.33)	13.54 - 179.86
Std Dev Myerson Centrality - Link detection algorithm (Method 3)			
Leisure network	185	9.07 (11.10)	0.50 - 58.53
Financial network	185	12.97 (12.45)	0.51 - 57.42
All network	185	9.71 (12.51)	0.55 - 55.90
Average Node Degree			
Leisure network	185	4.14 (1.65)	0.75 - 7.11
Financial network	185	4.31 (1.64)	1.38 - 7.14
All network	185	4.84 (1.68)	1.58 - 7.61
Std Dev Node Degree			
Leisure network	185	2.31 (0.60)	1.02 - 3.88
Financial network	185	2.55 (0.67)	1.18 - 4.33
All network	185	2.65 (0.68)	0.71 - 4.41

Note : The sample is restricted to villages with outside contact information and control variables.

Table 4. First stage effect at the Household level - Inside contact characteristics

	Observations $N_C + N_T$ [1]	**Control Mean [SD]** [2]	**Treatment Mean diff. (SE)** [3]
Household declaring inside contact			
Household having least one inside contact	N= 18642	0.86	-0.006
	9816C + 8826T	[0.35]	(0.011)
Number of inside contacts	N= 18642	2.60	-0.129**
	9816C + 8826T	[1.93]	(0.058)
Borrowing capacity			
Total emergency borrowing capacity - *Rs 1,000*	N= 18642	20.51	-2.078*
	9816C + 8826T	[41.72]	(1.185)
Total business borrowing capacity - *Rs 1,000*	N= 18642	24.08	-2.569*
	9816C + 8826T	[50.53]	(1.387)
Maximum total borrowing capacity - *Rs 1,000*	N= 18642	25.90	-2.457*
	9816C + 8826T	[52.15]	(1.421)
Total actual borrowed amount - *Rs 1,000*	N= 18642	6.96	-0.799**
	9816C + 8826T	[16.78]	(0.379)

Note : The sample is restricted to villages with outside contact information and at least 40 households. Borrowing capacity amounts have been top-coded using the 99^{th} percentile.. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively. Column (1) reports the total number of observation and its decomposition by groups. Column (2) reports the average outcome (standard deviation) for the control group. Column (3) reports the regression coefficent associated to the treatment dummy when controlling for pair fixed effects and with error terms cluster at the Service Area level.

TABLE 5. First stage effect at the Household level - Outside contact characteristics

	Observations $N_C + N_T$ [1]	Control Mean [SD] [2]	Treatment Mean diff. (SE) [3]
Household declaring outside contact			
Household having least one outside contact	N= 18642	0.52	-0.008
	9816C + 8826T	[0.50]	(0.012)
Number of outside contacts	N= 18642	0.90	-0.023
	9816C + 8826T	[1.15]	(0.029)
Borrowing capacity			
Total emergency borrowing capacity - *Rs 1,000*	N= 18642	31.44	-2.108
	9816C + 8826T	[73.34]	(1.449)
Total business borrowing capacity - *Rs 1,000*	N= 18642	35.14	-2.753*
	9816C + 8826T	[83.47]	(1.629)
Total actual borrowed amount - *Rs 1,000*	N= 18642	12.30	-1.154**
	9816C + 8826T	[32.38]	(0.537)
Maximum total borrowing capacity - *Rs 1,000*	N= 18642	37.22	-2.096
	9816C + 8826T	[86.92]	(1.709)

Note : The sample is restricted to villages with outside contact information and at least 40 households. Borrowing capacity amounts have been top-coded using the 99^{th} percentile.. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively. Column (1) reports the total number of observation and its decomposition by groups. Column (2) reports the average outcome (standard deviation) for the control group. Column (3) reports the regression coefficent associated to the treatment dummy when controlling for pair fixed effects and with error terms cluster at the Service Area level.

Table 6. First stage effect at the contact level - Outside contact characteristics

	Observations $N_C + N_T$ [1]	**Control Mean [SD]** [2]	**Treatment Mean diff. (SE)** [3]
Demographics			
Respondent's Age	N= 18103	41.59	-0.245
	9555C + 8548T	[12.15]	(0.239)
Male Respondent	N= 18103	0.44	-0.003
	9555C + 8548T	[0.50]	(0.008)
Contact's Age	N= 18101	43.27	-0.582***
	9554C + 8547T	[12.37]	(0.207)
Male Contact	N= 18103	0.67	-0.006
	9555C + 8548T	[0.47]	(0.007)
Type of contact			
Type of contact: Family and other relatives	N= 18103	0.70	-0.016
	9555C + 8548T	[0.46]	(0.011)
Type of contact: Employer	N= 18103	0.04	0.004
	9555C + 8548T	[0.19]	(0.005)
Nr. of days over the last 7 spent with the contact	N= 18103	2.86	0.126***
	9555C + 8548T	[2.50]	(0.039)
Location			
In the respondent's panchayat but not their gramum	N= 18103	0.02	0.005*
	9555C + 8548T	[0.12]	(0.003)
In the respondent's district but not their panchayat	N= 18103	0.66	0.034**
	9555C + 8548T	[0.47]	(0.014)
In Tamil Nadu but not the respondent's district	N= 18103	0.30	-0.038***
	9555C + 8548T	[0.46]	(0.013)
In India but not in Tamil Nadu	N= 18103	0.01	0.000
	9555C + 8548T	[0.09]	(0.001)
Outside of India	N= 18103	0.01	-0.002
	9555C + 8548T	[0.12]	(0.002)

Note : The sample is restricted to villages with outside contact information.. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively. Column (1) reports the total number of observation and its decomposition by groups. Column (2) reports the average outcome (standard deviation) for the control group. Column (3) reports the regression coefficent associated to the treatment dummy when controlling for pair fixed effects and with error terms cluster at the Service Area level.

A.3. **Household Income and Myerson Centrality.** Here we test if villagers' incomes are also positively correlated with their Myerson centralities, as predicted by the theory. The household income has been collected as part of the auxiliary survey in a subset of villages. The exact question was:

How much rupees, in total, did household members earn in the last 30 days from all income-generating activities including household business, farming, income from other sources of labour, transfers and government schemes? Include in-kind earnings, but first convert to cash and hen add to the total.

TABLE 7

	N [1]	**Mean (SD)** [2]	**Min-Max** [3]
Household Income over the last 30 days			
Monthly Income, *Rs*	8735	7603.17 (11293.11)	0.00 - 4.0e+05

Note : The sample is restricted to villages with outside contact information and control variables.

TABLE 8. Pearson Correlation Coefficient - Undirected networks, excluding Money Lenders

	Income over the last 30 days [1]
Myerson Centrality - Incoming link deletion algorithm (Method 1)	
Leisure network	0.037 ***
Financial network	0.034 ***
Financial & leisure network	0.027 ***
Myerson Centrality - Outgoing link detection algorithm (Method 2)	
Leisure network	0.032 ***
Financial network	0.030 ***
Financial & leisure network	0.023 ***
Myerson Centrality - Link detection algorithm (Method 3)	
Leisure network	0.048 ***
Financial network	0.043 ***
Financial & leisure network	0.050 ***
Node Degree	
Leisure network	0.060 ***
Financial network	0.089 ***
Financial & leisure network	0.081 ***

Note : The sample is restricted to villages with outside contact information and control variables.

Appendix B. Variable Construction

B.1. **Approximating the Myerson distance and centrality.** Assuming that risk sharing results in agents share the surplus generated by information risk sharing according to the Myerson value, Ambrus and Elliott (2018) show that payoffs can be calculated by applying the inclusion–exclusion principle from combinatorics, and relate this to a measure of the distance between agents on the risk sharing agent. They call this distance measure Myerson distance. We would like to compute the Myerson distance of every pair in every village and the Myerson centrality for all nodes. Unfortunately, this is computationally infeasible for the sample sizes of our data (see Algaba et al. (2007)), presenting a new challenge. Thus, we develop an approximation, described below.

Let $\mathbf{md}(L)$ be the matrix of Myerson distances and define $\mathbf{q}(L) := 1/2 - \mathbf{md}(L)$. So $\mathbf{q}(L)$ is a matrix with the ijth entry capturing the probability that, upon his arrival agent i will not be connected to agent j. It is difficult to directly characterize $\mathbf{md}(L)$ (or equivalently, $\mathbf{q}(L)$) as each village typically consists of around 230 households and the number of candidate paths between each i and j is exponential in the size of the network. Correctly accounting for paths that share nodes is computationally very intensive, and it has to be done for all pairs of agents without a direct connection.[11] Instead, we develop a computationally feasible approximation of $\mathbf{md}(L)$, which is exact for trees.

To approximate $\mathbf{q}$, we use the following idea. The algorithm works by starting with a node, moving to its neighbors, then move to its neighbors' neighbors, and so on, never returning to a previously used node along a given walk. This helps us to avoid counting walks that revisit nodes and are therefore not paths. All the while, we keep track of how many ways we have moved from the original node to any given node. We denote our approximation of $\mathbf{q}$ by $\widehat{\mathbf{q}}$.

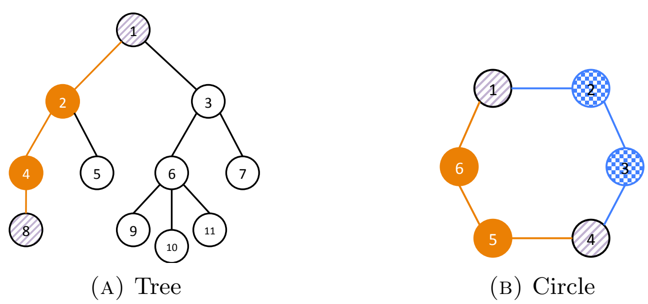


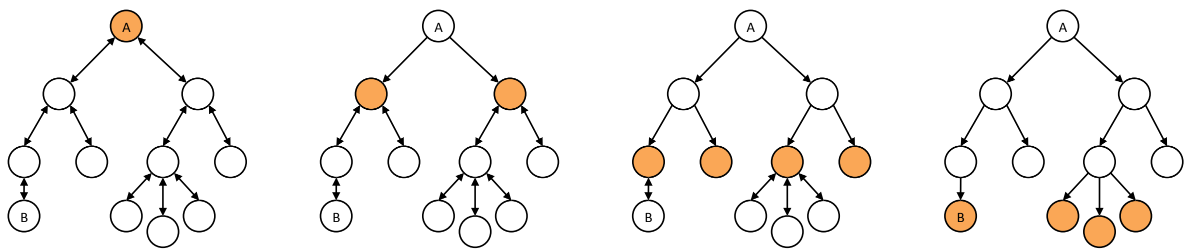
Figure 1. The nodes i,j for which we are computing $md(i, j, L)$ have purple stripes. The tree contains a single path (solid orange nodes), whereas the circle contains two paths (solid orange nodes and chequered blue nodes).

The inclusion–exclusion principle weights paths that are longer less and a path that shares many nodes with another less. With this in mind, we choose the following two approximation

[11]Further, due to presumed measurement error (see Banerjee et al. (2013)), there are likely to be missing paths. In fact, the data have occasional disconnected components, and so measures that are precisely based on exact paths or even maximal path lengths are likely to be problematic (Chandrasekhar and Lewis (2014)).

strategies. Let the shortest path between two nodes be of length l. We first count the paths of length l and length $l+1$. We then count paths of length $l+2$.[12] If there are fewer than k such paths, we use them all. Otherwise, we consider only the k shortest and in practice we set $k = 4$.[13] Discarding longer paths in this way biases downwards our approximation of $\mathbf{q}$. As we cannot keep track of exactly which nodes feature in each path, we also have to make an assumption about the overlap of nodes in order to apply the inclusion–exclusion principle to these paths. Each path must share the same first and last node. We perform the inclusion–exclusion principle assuming that only these nodes are shared. Assuming no other nodes are shared introduces a second bias, but this time upwards in our approximation.

To explain these concepts, we provide some illustrations. Figure 1 presents two examples: a tree and a circle. The tree has a single path between nodes 1 and 8, whereas the circle has two paths between nodes 1 and 4. Figure 2 shows how links are removed for the case of a tree. Once a node has been reached, links back into that node are deleted before the nodes neighbors are "infected." This ensures only paths, and not other walks, are included in the calculation.



(A) (B) (C) (D)

FIGURE 2. As the algorithm progresses, directed links into nodes that are reached are deleted. This ensures that only paths, and not other walks, are included. In this case, as in all tree networks, there is a unique path from A to B.

In the case of the circle shown in Figure 1, our algorithm is also exact for paths between 1 and 4. There are two paths (which in this case are both shortest paths too), and we find both in the initial run of our algorithm. Following the inclusion–exclusion principle, we add 1/4 to 1/4 and subtract 1/6. In this case our assumption that the two paths share only two nodes is accurate. We are also exact for paths between 1 and 3, but in this case there is a path of length $l+2$. To find this path, we look for paths of length l from 1 to nodes other than 3. In this case there is one such path to node 5. We then look for paths from 5 to 3 that pass through one other node. There is one such path and so the calculation we perform

[12]Counting more paths greatly (exponentially) increases the running time of our algorithm.

[13]We need a fixed (small) truncation. Otherwise both the memory requirements and the run-time of the algorithm grow exponentially. Results are not sensitive to the truncation point.

is: $1/3+1/5-1/6$. While we are accurate for all pairs of nodes in the circle shown, in larger circles we will miss the longer paths.

The following algorithm finds the length of the shortest path between two nodes, how many paths of that length there are and how many paths there are that are one longer. From this information, we also find paths of length $l+2$.[14]

Algorithm 1 (Incoming Link Deletion)**.** *Let* e^i *be the ith basis vector. This will represent the root (starting) node. Initialize* $\widehat{\mathbf{q}} = zeros(n,n)$*, a matrix of zeros. Initialize* $z^{t,i} = zeros(n,1)$ *and* $x^{t,i} = zeros(n,1)$ *to be n-vectors of zeros, indexed by* $i = 1,...,n$ *and* $t = 1,...,T$*. Repeat steps 1–4 for each of* $(e^1,...,e^n)$*.*

(1) *Period 1: There is no identification or updating steps.*
   (a) *Percolation:* $x^{1,i} = \mathbf{A}e^i$.
   (Identifies who is connected to the root node)
(2) *Period 2, given* $(x^{1,i}, \mathbf{A})$*:*
   (a) *Identification:* $z^{2,i} = e^i$.
   (b) *Update graph:*[15] $\mathbf{A}_2 = zeros(n,n),\ \mathbf{A}_2(\neg z^{2,i},:) = \mathbf{A}(\neg z^{2,i},:)$.
   (Deletes links into the root node)
   (c) *Percolation:* $x^{2,i} = \mathbf{A}_2 x^{1,i}$.
   (Records number of paths from root node to other nodes passing through one other)
(3) *Period t, given* $(x^{t-1,i}, \mathbf{A}_{t-1})$*:*
   (a) *Identification:* $z^{t,i} = \mathbf{1}\left\{\sum_{s=3}^{t} x^{s-2,i} > 0\right\}$.
   (Identifies nodes already visited)
   (b) *Update graph:* $\mathbf{A}_t = zeros(n,n),\ \mathbf{A}_t(\neg z^{t,i},:) = \mathbf{A}_{t-1}(\neg z^{t,i},:)$.
   (Deletes links into all nodes that have already been visited)
   (c) *Percolation:* $x^{t,i} = \mathbf{A}_t x^{t-1,i}$.

By construction $x_j^{t,i}$, the jth entry of $x^{t,i}$, records paths from i to j that pass through t nodes. If t' is the lowest t with a positive entry in this matrix, then the shortest path from i to j passes through t' nodes. In this case, $x_j^{t',i}$ tells us how many such paths there are and $x_j^{t'+1,i}$ tells us how many paths there are that pass through one more node. However, by construction $x_j^{t'+k,i} = 0$ for all $k > 1$ and longer paths are not recorded. This is because the incoming links to node j will have been deleted by this step of the algorithm. Deletion of incoming links helps prevents walks that are not paths from being recorded. Using this information for all seed nodes, the number of paths of length $t'+2$ between i and j are also found as described above. The inclusion-exclusion principle is then applied to this combined set of paths, assuming each path shares only the first and last nodes, to calculate $\widehat{\mathbf{q}}(L)$.

[14]For paths from i to j, this is done by looking at paths of length l to agents other than j, and then looking at paths from these agents to j.
[15]Let $\mathbf{A}(:,v)$ denote $(A(1,j),...,A(n,j))$.

Ambrus and Elliott (2018) relate a measure of agents centrality in the network to their incentives to form out-of-group links, and term this centrality measure Myerson centrality. To approximate Myerson centrality we use $\widehat{MC}_i = \sum_j \widehat{q}_{ij}$. While this approximation generates a cardinal measure of what is an ordinal concept, it does correctly order people when the Myerson distance approximation is exact as shown in Proposition 2

Proposition 2. *If i is more Myerson central than j, then $\sum_k q_{ik} > \sum_k q_{jk}$.*

Proof. By definition, if i is more Myerson central than j, then there exists a pairing or arrival orders, such that for each arrival order in which j is path-connected to k agents, i is path-connected to weakly more than k agents. Thus, if i is more Myerson central than j, then the expected number of agents that i is connected to upon her arrival is greater than the expected number of agents that j is connected to upon her arrival. The expected number of agents that i is connected to upon her arrival is $\sum_k q_{ik}$. We therefore have that $\sum_k \widehat{q}_{ik} > \sum_k \widehat{q}_{jk}$ as claimed. □

We now show that the Myerson distance approximation is exact for trees.

Proposition 3. *Let L be a tree. Then $\widehat{\mathbf{q}}(L) = \mathbf{q}(L)$.*

Proof. We will say that agent k is a distance-t neighbor of i if the shortest path from i to k take exactly t steps (and contain $t+1$ agents, including i and k).

Consider the implementation of the Incoming Link Deletion algorithm to find $\widehat{q}_{ij}$. We begin by calculating $x^{1,i} = \mathbf{A}e^i$, where e^i is the ith basis vector. This identifies all agents connected to i. We then set all entries in the ith row from the adjacency matrix $\mathbf{A}$ to 0 and call this new matrix $\mathbf{A}_2$. This deletes the inward links to$i$ in the network L. Starting from i's neighbors, we then find their neighbors on $\mathbf{A}_2$. In other words we calculate $x^{2,i} = \mathbf{A}_2 x^{1,i}$. This identifies the distance-2 neighbors of i. We then delete the rows of $\mathbf{A}_2$ that are indexed by one of i's neighbors, and so on.

In the tth round the algorithm identifies the distance-t neighbors of i. Thus, for $t < l$, $x_j^{l,i} = 0$; for $t = l$, $x_j^{l,i} = 1$; and for all $t > l$, $x_j^{t,i} = 0$. Deleting incoming links ensures for all $t > l+1$, $x_j^{t,i} = 0$. As L is a tree there, there is no path of length $l+1$ to j and so $x_j^{l+1,i} = 0$. The algorithm therefore finds the unique path from any i to any j and records its length; If the unique path from i to j has length l, $\widehat{q}_{ij} = 1/l$. From equation (11) in Ambrus and Elliott (2018) it is then easily verified that $q_{ij} = 1/l$. Thus $\widehat{\mathbf{q}}(L) = \mathbf{q}(L)$. □

In combination, Propositions 2 and 3 lead directly to the following corollary.

Corollary 4. *Let L be a tree. If i is more Myerson central than j, then $\widehat{MC}_i > \widehat{MC}_j$.*

A limitation of the Incoming Link Deletion algorithm is that longer paths are excluded. To address this, we construct an alternative algorithm. This Outgoing Link Deletion algorithm is identical to the one described, except that it deletes outgoing links instead of incoming links. The Outgoing Link Deletion algorithm finds longer paths, and does an especially good job of picking up longer paths that share few nodes with other paths. However, it also includes additional short walks that are not paths and is not exact for tree networks. As longer paths are found, we directly use the output of the algorithm without constructing any additional longer paths. Nevertheless, for the set of paths we find, it is computationally infeasible to compute the Myerson distances using the inclusion-exclusion principle. Censoring these paths would defeat the point of the Outgoing Link Deletion algorithm. Instead, we use an approximation of the inclusion–exclusion principle which makes the computation much simpler. This approximation treats every path as completely independent, assuming that no nodes are shared (even though we know at two must be). For example, if we find 3 paths from i to j that pass through l nodes, l' nodes and l'' nodes respectively, our approximation of q_{ij} will be $1/l + 1/l' + 1/l''$. Finally, we also consider a hybrid of the Incoming Link Deletion algorithm and the Outgoing Link Deletion algorithm. We refer to this as the Link Deletion algorithm.