# Development Economics

## Networks

Emily Breza, Harvard University

AEA Continuing Education

January 2024

# Importance of Networks for Development

Developing countries face market incompleteness

- $\implies$ Reliance on informal institutions to fill the gap

Important for numerous domains:

- Financial: risk sharing, credit (monitoring and screening)
  - Have already seen numerous examples in the context of social transfers/insurance
- Information: job referrals, technology adoption, access to new government programs, advice, aspirations
- Social: religious events, festivals, sports,...

# Roadmap

1. Value of Networks
2. Introduction to Networks
3. Information Diffusion and Aggregation
4. Network - Market Interactions

# Cai and Szeidl (2018) QJE

Question: What is the value of a firm's network?

- Potential benefits: information, introductions to customers/suppliers, contracting relationships, trade credit, collusion...

Design: experiment to change the networks (very difficult!)

- Managers 2,800 of SMEs in Nanchang, China
- Create groups, encourage to sustain self-enforced monthly meetings
- Government involvement helps here – use certificate as incentive

Design details

- Half of firms in meetings treatment arm
- Meetings firms randomized into groups of 10
- Additional treatments to explore mechanisms

# Cai and Szeidl (2018): Large RF impacts!

FS: main treatment $\implies \uparrow$ direct and indirect relationships.

Table 3: Effect of Meetings on Firm Performance

| Dependent var.: | log Sales | Profit (10,000 RMB) | log Number of Employees | log Total Assets | log Productivity | log Reported - log Book Sales |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Post | 0.00533 | 8.6879* | 0.0176 | 0.0170 | 0.0152 | 0.0004 |
| (1=Yes, 0=No) | (0.0198) | (4.5078) | (0.0166) | (0.0191) | (0.0217) | (0.0071) |
| Meetings*Post | 0.0749** | 21.6519** | 0.0524** | 0.0530 | 0.0675* | 0.0037 |
| | (0.0361) | (10.5511) | (0.0264) | (0.0346) | (0.0392) | (0.012) |
| Observations | 5,292 | 5206 | 5,292 | 5,292 | 5126 | 5220 |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.004 | 0.009 | 0.006 | 0.003 | 0.004 | 0.0001 |

Note: Standard errors clustered to the meeting group level for treated firms and to the firm level for control firms. Productivity is measured by the ratio between valueadded and number of employee. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Effect of Meetings on Intermediate Outcomes

| Dependent var.: | log Number of Clients | log Number of Suppliers | Bank Loan | Informal Loan | Tax/Sales | Stress |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Post | 0.0142 | 0.0245 | -0.0396*** | 0.0905*** | 0.000593 | 0.00531 |
| (1=Yes, 0=No) | (0.0201) | (0.0218) | (0.0108) | (0.0113) | (0.000976) | (0.0195) |
| Meetings*Post | 0.0894*** | 0.0811*** | 0.0907*** | 0.0521*** | 0.000728 | 0.0448 |
| | (0.0298) | (0.0314) | (0.0156) | (0.0175) | (0.00149) | (0.0277) |
| Observations | 5,280 | 5,182 | 5,292 | 5,292 | 5,292 | 5,292 |
| Firm FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.010 | 0.010 | 0.013 | 0.073 | 0.001 | 0.003 |

Note: Standard errors clustered to the meeting group level for treated firms and to the firm level for control firms. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# Cai and Szeidl (2018): Mechanisms

Information seeded about a valuable, competitive grant (worth $\$32,000$). Fraction receiving info $\in \{0, 0.5, 0.8\}$

| Dependent var.: | Applied for the Firm Funding Product | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Sample:* | *All Firms* | | *Uninformed Firms in Meetings* | | |
| Info | 0.300*** | 0.370*** | | | |
| | (0.0208) | (0.0227) | | | |
| No Info * Meetings | | 0.202*** | | | |
| | | (0.0247) | | | |
| Info * Meetings | | 0.0721** | | | |
| | | (0.0323) | | | |
| Having Informed Group Members | | | 0.315*** | | 0.402*** |
| | | | (0.0340) | | (0.0470) |
| Competition | | | | -0.155*** | -0.0715** |
| | | | | (0.0497) | (0.0344) |
| Having Informed Group Members *Competition | | | | | -0.173*** |
| | | | | | (0.0605) |
| Firm Demographics | No | No | Yes | Yes | Yes |
| Observations | 2,646 | 2,646 | 846 | 846 | 846 |
| R-squared | 0.114 | 0.148 | 0.140 | 0.111 | 0.242 |

- Large information spillovers!
- Notice less so when firms are competitors

# Roadmap

1. Value of Networks
2. Introduction to Networks
3. Information Diffusion and Aggregation
4. Network - Market Interactions

## Representing Networks

- $V = \{1, .., n\}$ - a set of vertices/nodes/agents
- $E$ - a set of edges
- $A$ - adjacency matrix, $a_{ij} \in \{0, 1\} \Leftrightarrow ij \in E$ - encodes edge

Networks are complex

- Suppose 20 nodes. How many possible graphs $A$?
- Person 1 can have 19 links, person 2 can have 18, etc

$$\binom{20}{2} = 190$$

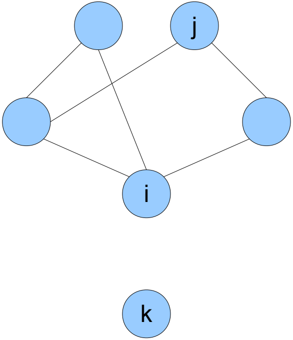
- Each link present or not

$$2 \times ... \times 2 = 2^{\binom{n}{2}} = 2^{190}$$

- number of atoms in universe: around $2^{240}$

Need to reduce dimensionality to make progress

# Path Length: Social Distance




- *$Path_{ij}$* sequence of connected nodes from $i$ to $j$, nodes distinct
- *$SocialDistance_{ij}$* is the shortest path from node $i$ to $j$
- Node $k$ is *unreachable* by any other node
- The *giant component* contains all nodes other than $k$
- Diameter: longest shortest path (here 2)

# Centrality

Many measures, including:

- *Degree*: number of links a node has

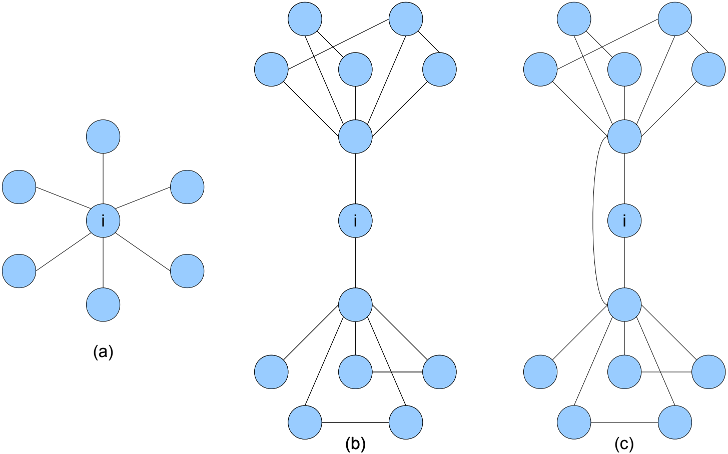
$$d = A \cdot 1 = \left( \sum_j a_{ij} \right)_{i=1}^{n}$$

- *Eigenvector Centrality*:

$$\lambda C_i^e(A) = \sum_j a_{ij} C_j^e(A)$$
$$\lambda C^e(A) = AC^e(A)$$

- *Betweenness Centrality*: Fraction of shortest paths between all other nodes a given node belongs to.

# Centrality




- Centrality measures need not overlap
- Empirically, tend to be correlated but still distinct

# Properties of Real World Social Networks

- Small worlds: small diameters (longest shortest paths) and small average path lengths
- High clustering coefficients, relative to links being generated independently at random (10,000 times more in some applications!)
  - Friends of friends are typically also directly connected. (Triangles in network)
- Very large giant component (most people are connected in some way, directly or indirectly)
- Fat-tailed degree distribution (small number of people have extremely large number of friends)
- Homophily (either by opportunity or choice)

# Roadmap

1. Value of Networks
2. Introduction to Networks
3. Information Diffusion and Aggregation
4. Network - Market Interactions

# Technology Adoption

Social learning has long been studied to understand technology adoption:

- Planting decisions and harvests observable to neighbors
- Active information networks among local farmers
- In many contexts, top-down policies can't explain adoption patterns (gov't policies often not very strong)

# S-Shaped Technology Adoption: Drug Prescriptions, Hybrid Corn

Coleman et al. ('66), Griliches ('57)

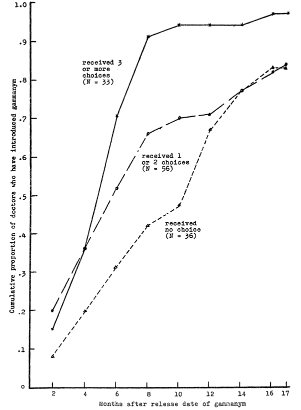
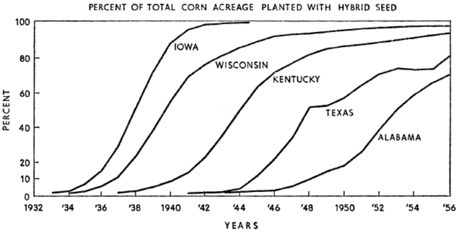



Fig. 2. Cumulative proportion of doctors introducing gammanym: differen




S-shaped adoption can arise from peer spillovers!

# Learning: Diffusion

Goal: can we understand how information about new technologies spreads through the network?

- Q1: Who to target?
- Q2: Aside from information effects, are there endorsement effects?

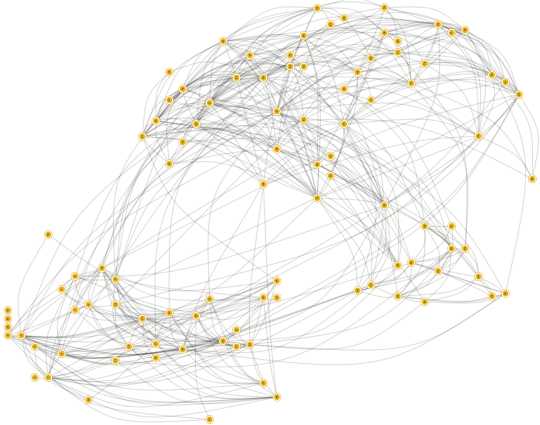
Banerjee et al (2013) take first pass as this question in economics. Quasi-experimental variation to investigate:

- Application of technology adoption to microfinance – who adopts?
- Agents need to be aware of MF, decide on suitibility
  - (not obvious best application due to group structure etc.)

Design: differences-in-differences

- 75 villages with network surveys
- MFI entered some but not all
- Fixed strategy for who to inform first "injection points", induces variation in network characteristics

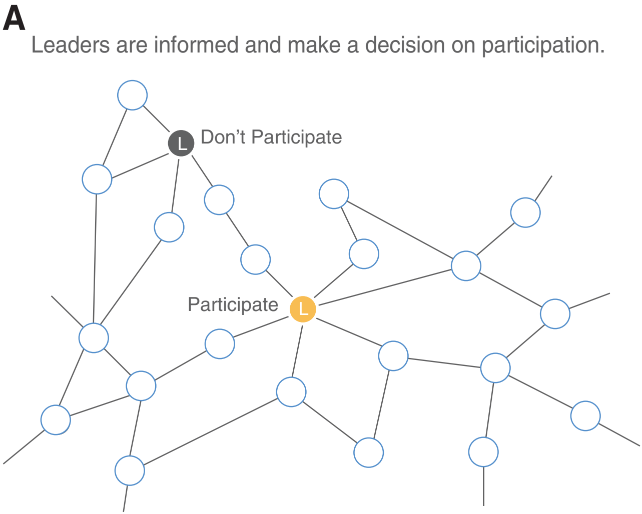
# Aside: Karnataka Village network data



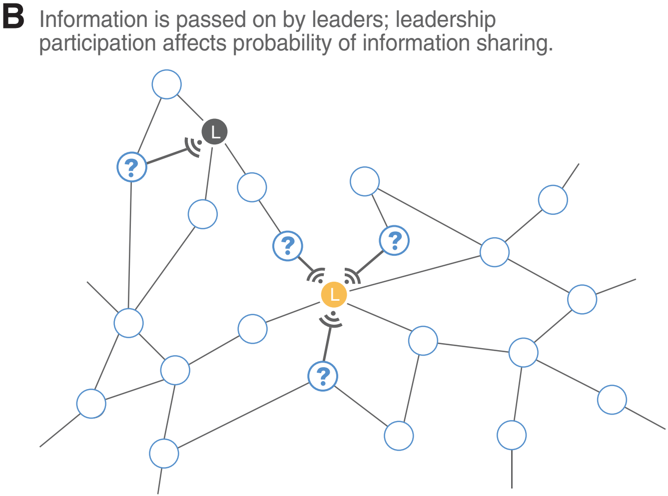
Panel data (2 waves):

- Relationships: relatives, friends, creditors, debtors, advisors and religious company
- Often, use undirected, unweighted OR network
- Basic demographics: caste, GPS, occupation, ...
- Typically only feasible when n small (i.e., villages)

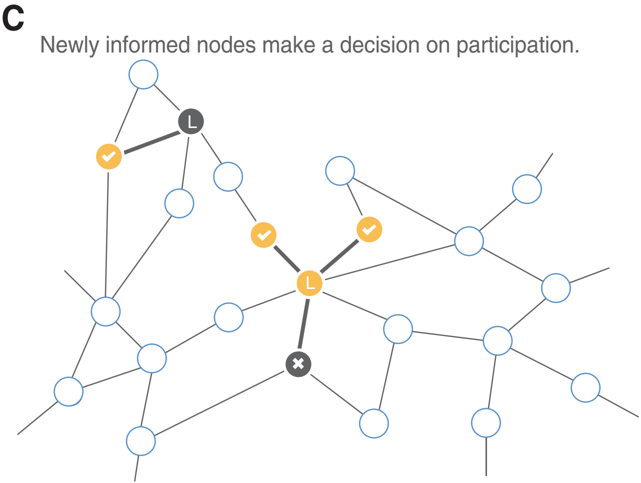
# Diffusion of Microfinance



# Diffusion of Microfinance

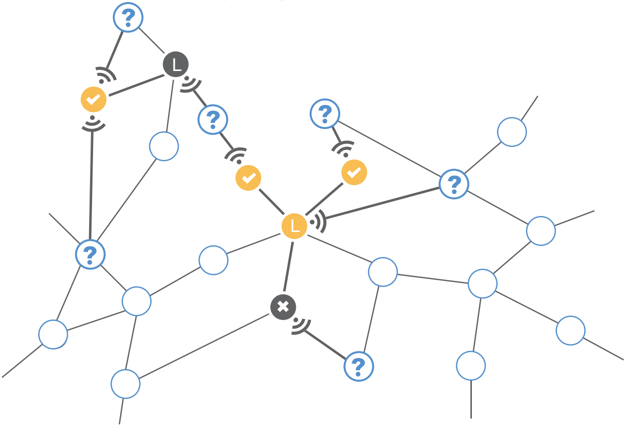


# Diffusion of Microfinance

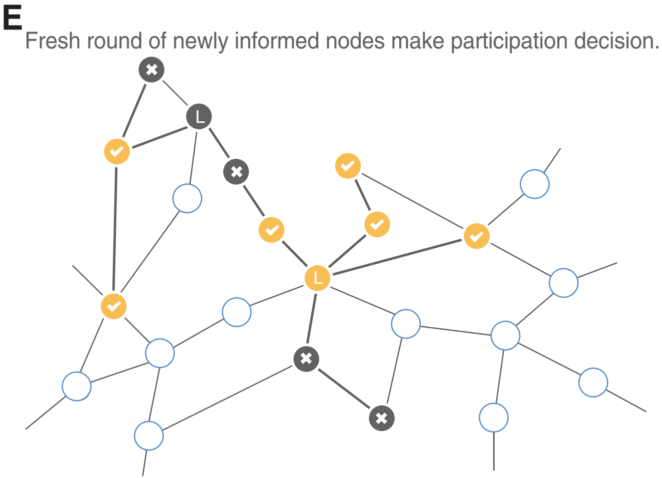


# Diffusion of Microfinance

**D** All informed nodes pass on information further; the probability of information sharing is, again, based on participation.



# Diffusion of Microfinance




What happens if process keeps going for T large?

# Diffusion of Microfinance: Where to inject?

Policy-relevant question: where to inject?

- Central agents - more influential
- But what measure of centrality?

New measure: diffusion centrality

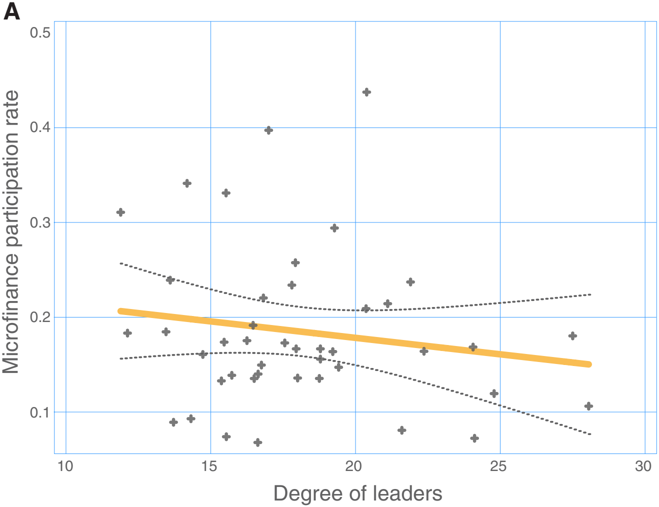
- Hearing matrix H's ijth element gives the expected number of times j hears about info originating from i.
- $DC_i$ gives the expected number of times all nodes taken together hear the message originating from i
  - More times $\uparrow$ likelihood of remembering, details learned etc.
  - Different from simple, viral diffusion (Akbarpour et al 2020)

$$H(A; q, T) = \sum_{t=1}^{T} (Aq)^t$$

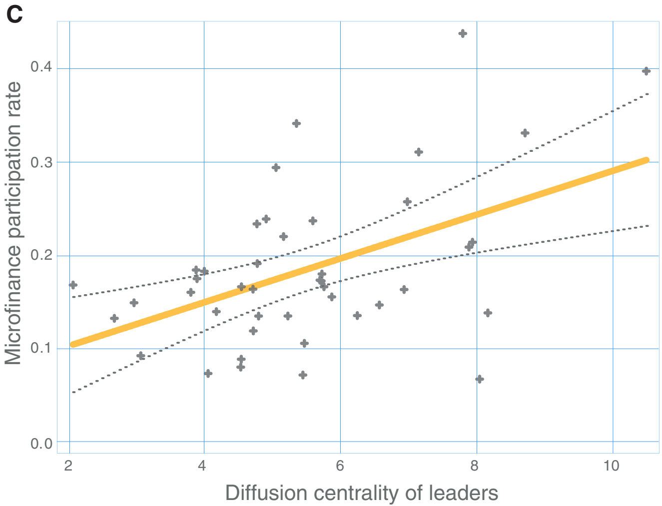
$$DC(A; q, T) = H(A; q, T) \cdot 1$$

What measure works better in the data?

# Diffusion of Microfinance: Village-Level Take-Up and Centrality



# Diffusion of Microfinance: Village-Level Take-Up and Centrality



# Is this useful?

Initial policy reaction: How to use?

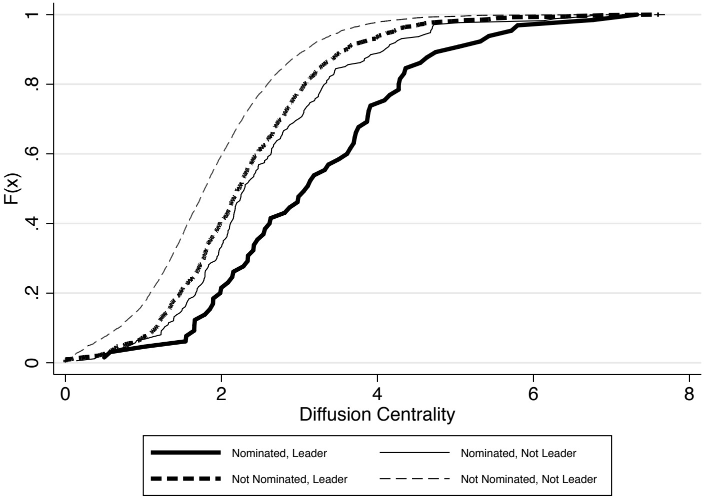
- Network data is expensive, doesn't seem practical.

How about asking a few people in the network? Same team of researchers tries the following:

**Eliciting centrality**

1. *"If we want to spread information about a new loan product to everyone in your village to whom do you suggest we speak?"*
2. *"If we want to spread information to everyone in the village about tickets to a music event, drama, or fair that we would like to organize in your village, to whom should we speak?"*

# More central, more nominations: Event



# Experimental Validations

Experiment: spread of immunization in Haryana

- 516 villages were seeded information on immunization
- random, Trusted, "Gossip" or Trusted Gossip.
- Gossip increase number of kids immunized for all different shots by 20%

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Penta1 level | Penta2 level | Penta3 level | Measles1 level | Number of Children |
| | (1) | (2) | (3) | (4) | (5) |
| gossip | 1.017* | 1.022* | 1.030** | 1.078** | 4.903* |
| | (0.603) | (0.561) | (0.523) | (0.500) | (2.503) |
| trusted | 0.261 | 0.302 | 0.490 | 0.439 | 1.849 |
| | (0.486) | (0.448) | (0.418) | (0.408) | (2.047) |
| trustgossip | 0.479 | 0.526 | 0.514 | 0.444 | 2.376 |
| | (0.470) | (0.429) | (0.396) | (0.376) | (1.917) |
| Observations | 6697 | 6697 | 6697 | 6697 | 6712 |
| Villages. | 521 | 521 | 521 | 521 | 521 |
| Mean (Random Seeds) | 4.31 | 4.06 | 3.71 | 3.53 | 18.11 |
| Gossip=Random (pval.) | 0.092 | 0.069 | 0.049 | 0.032 | 0.051 |
| Gossip=Trusted (pval.) | 0.176 | 0.168 | 0.268 | 0.182 | 0.192 |
| Gossip=Trusted Gossip (pval.) | 0.343 | 0.338 | 0.281 | 0.166 | 0.271 |

## Application: Savings and Reputation

In theories of MF/ROSCAs,"social reputation" often assumed

> *"the contributing member may admonish his partner for causing him or her discomfort and material loss. He might also report this behavior to others in the village, thus augmenting the admonishment felt. Such behavior is typical of the close-knit communities in some LDCs."*
>
> *– Besley and Coate (1995)*

But challenging to identify inner workings of ROSCAs/MFIs

- Strategic game with many members!

Breza and Chandrasekhar (2019) Econometrica Approach: simplify the problem, use insights from network theory

- RCT of stylized savings intervention
- Recruit individuals who want to save more, have capacity
- Give everybody bank account, reminders, goal setting
- Randomize addition of monitor: peer in village who sees savings progress

# Breza and Chandrasekhar (2019)

- Result 1: Randomly assigned monitor ↑ savings by 35%, improvements in shock mitigation
- But which kinds of monitors drive results? Model building off of ideas of diffusion centrality
- Reputation cost from not reaching goal. Monitor most effective if:
  - Many people learn (centrality)
  - Those who learn are likely to be relevant for saver (distance)

| *Dependent Variable* | (1) Log Total Savings | (2) Log Total Savings | (3) Log Total Savings | (4) Log Total Savings | (5) Log Total Savings | (6) Log Total Savings |
|---|---|---|---|---|---|---|
| Monitor Centrality | 0.178** | | 0.134* | | 0.153** | |
| | (0.0736) | | (0.0729) | | (0.0675) | |
| Saver-Monitor Proximity | | 1.032*** | 0.865** | | 1.108*** | |
| | | (0.352) | (0.334) | | (0.294) | |
| Model-Based Regressor | | | | 1.450** | | 1.819*** |
| | | | | (0.693) | | (0.632) |
| R-squared | 0.150 | 0.155 | 0.161 | 0.148 | 0.101 | 0.080 |
| Fixed Effects | Village | Village | Village | Village | | |
| Controls | Saver, Monitor | Saver, Monitor | Saver, Monitor | Saver, Monitor | Double-Post LASSO | Double-Post LASSO |

# Breza and Chandrasekhar (2019)

Can also ask whether reputations change as function of treatment.

- 560+ random respondents chosen 15 mo. after end of intervention, asked about 8 participants
- asked if each saver was responsible, good at meeting goals
- is respondent more likely to say "Yes" when the saver truly did meet her savings goal (or "No" when the saver didn't) when the random monitor is more central?

| | (4) | (5) | (6) |
|---|---|---|---|
| *Dependent Variable: Beliefs about Saver* | Good at Meeting Goals | Good at Meeting Goals | Good at Meeting Goals |
| Monitor Centrality | 0.0389 | 0.0374 | 0.0353 |
| | (0.0144) | (0.0140) | (0.0148) |
| Respondent-Monitor Proximity | 0.0476 | 0.0181 | 0.0360 |
| | (0.0422) | (0.0366) | (0.0342) |
| Observations | 4,743 | 4,743 | 4,743 |
| R-squared | 0.030 | 0.023 | 0.314 |
| Fixed Effects | No | Village | Respondent |
| Controls | Saver | Saver | Saver |

Central monitor causes beliefs to be updated in direction of actual goal attainment (13.3%)

# Beaman et al 2021 AER: Diffusion of agri. technique

"Can Network Theory-based Targeting Increase Technology Adoption'?' by Lori Beaman, Ariel BenYishay, Jeremy Magruder, Mushfiq Mobarak

- Question: how to seed information about a new technology
  - Focus on on simple vs. complex contagion

- Simple contagion: 'viral' infection, only need to hear once
- Complex contagion: need to hear multiple times (may forget or may need to aggregate different signal draws etc.)
  - Specify a threshold model: only adopt if a threshold number of neighbors also adopts
  - CC on adoption, not simply hearing info

# Beaman et al 2021

- Setting
  - 200 villages in Malawi (+ network data)
  - interested in diffusion of pit planting technique
  - authors calculate experimental returns to adoption – 40%!

# Beaman et al 2021

- Setting
  - 200 villages in Malawi (+ network data)
  - interested in diffusion of pit planting technique
  - authors calculate experimental returns to adoption – 40%!
- Design: 4 Treatments
  - Geographic (T1): seed info with geographically central
  - Extension (T2): seed info with extension worker's choice (status quo)
  - Network (T3 & T4): seed "optimally" from network under simple or complex contagion models
    - Simulate from the models before running the experiment T=4.
  - Essential: Can locate shadow seeds in each village: counterfactual seedings

# Simple vs. Complex Contagion

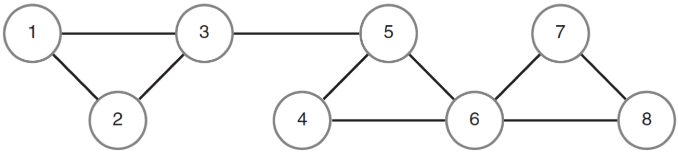



FIGURE 1. AN EXAMPLE NETWORK

Suppose extension officer can inform 2 people. Learning occurs over 3 periods:

- Who to target under simple contagion? (ie., only need to hear once)

# Simple vs. Complex Contagion

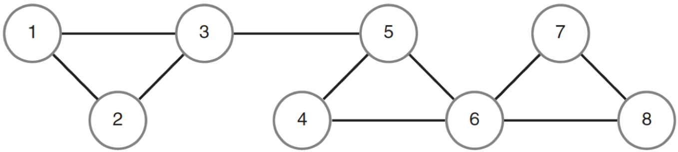



FIGURE 1. AN EXAMPLE NETWORK

Suppose extension officer can inform 2 people. Learning occurs over 3 periods:

- Who to target under simple contagion? (ie., only need to hear once)
- Who to target under complex contagion? (i.e., need to hear from 2 people) What is maximum number of non-seed farmers who adopt?

| | Any non-seed adopters | | Adoption rate | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Complex diffusion treatment | 0.252 | 0.304 | 0.036 | 0.036 |
| | (0.093) | (0.101) | (0.016) | (0.026) |
| Simple diffusion treatment | 0.155 | 0.189 | 0.036 | 0.006 |
| | (0.100) | (0.111) | (0.017) | (0.022) |
| Geographic treatment | 0.107 | 0.188 | 0.038 | 0.013 |
| | (0.096) | (0.110) | (0.027) | (0.034) |
| Year | 2 | 3 | 2 | 3 |
| Observations | 200 | 141 | 200 | 141 |
| Mean of Benchmark treatment (omitted category) | 0.420 | 0.543 | 0.038 | 0.075 |
| SD of Benchmark | 0.499 | 0.505 | 0.073 | 0.109 |
| *p-values for equality in coefficients* | | | | |
| Simple = Complex | 0.300 | 0.240 | 0.981 | 0.173 |
| Complex = Geo | 0.102 | 0.220 | 0.937 | 0.491 |
| Simple = Geo | 0.623 | 0.990 | 0.950 | 0.783 |

*Notes:* The reference group is the Benchmark treatment. The sample for year 3 (columns 2 and 4) excludes Nkhotakota district. The *Any non-seed adopters* indicator in columns 1–2 excludes seed farmers. The adoption rate in columns 3–4 include all randomly sampled farmers, excluding seed and shadow farmers. All columns include

| | Heard of pit planting | | | Knows how to pit planting | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Connected to 1 seed | 0.002 | 0.030 | 0.016 | 0.017 | 0.021 | −0.031 |
| | (0.024) | (0.022) | (0.029) | (0.016) | (0.017) | (0.023) |
| Connected to 2 seeds | 0.084 | 0.124 | 0.064 | 0.062 | 0.068 | 0.110 |
| | (0.038) | (0.040) | (0.064) | (0.028) | (0.029) | (0.051) |
| Within path length 2 of at least one seed | −0.018 | 0.016 | 0.067 | 0.005 | 0.022 | 0.028 |
| | (0.028) | (0.027) | (0.042) | (0.018) | (0.021) | (0.028) |
| Year | 1 | 2 | 3 | 1 | 2 | 3 |
| Observations | 4,155 | 4,532 | 3,103 | 4,155 | 4,532 | 3,103 |
| Mean of reference group (no connection to any seed) | 0.223 | 0.286 | 0.391 | 0.057 | 0.095 | 0.147 |
| SD of reference group | 0.416 | 0.452 | 0.488 | 0.232 | 0.293 | 0.355 |
| $p$-value for 2 connections = 1 connection | 0.018 | 0.013 | 0.442 | 0.072 | 0.091 | 0.004 |

*Notes:* Sample excludes seed and shadow farmers. The reference group is comprised of individuals with no direct or 2-path-length connections to a seed farmer. Only connections to simple, complex, and geo seed farmers are considered (no connections to Benchmark farmers included). The dependent variable in columns 1–3 is an indicator for whether the respondent reported being aware of a plot preparation method other than ridging and then subsequently indicated awareness of pit planting in particular. In columns 4–6, the dependent variable is an indicator for whether the farmer reported knowing how to implement pit planting. In all columns, additional controls include indicators for the respondent being connected to: one Simple partner, two Simple partners, one Complex partner, two Complex partners, one Geo partner, two Geo partners, within 2 path length of a Simple partner, within 2 path length of a Complex Partner, and within 2 path length of the geo partner. Also included are village fixed effects. Standard errors are clustered at the village level.

# Targeting to spread information and change norms

I study targeting of peer health intervention in Brazilian high schools with Erick Baumgartner, Eliana La Ferrara, Victor Orozco, and Pedro Rosa Dias

- Goal: improve information about contraception use, change norms around sexual health, increase protective behaviors
- Intervention delivered through peer volunteers ("mobilizers")
- Context of high teen pregnancy rates (10%), low contraceptive use & knowledge, limited communication about sex

Three methods to select peer educators

T1. Selection by school (status quo)

T2. Network centrality
  - Have the most reach in the network to **transmit information**

T3. Most popular students
  - "Social referents" may be best suited to **shape norms**

T4. Control group

# First stage outcomes (by treatment arm)

| VARIABLES | Knows the Teenager Booklet (1) | (2) | Received sexual health counseling in school (3) | (4) | (Count) friends I speak with about sexuality (5) | (6) |
|---|---|---|---|---|---|---|
| T1: school selects (based on 2018) | 0.039 | 0.037 | 0.034 | 0.025 | 0.086* | 0.073 |
| | (0.025) | (0.025) | (0.023) | (0.024) | (0.046) | (0.045) |
| T2: network centrality (based on 2018) | 0.069*** | 0.071*** | 0.075*** | 0.066*** | 0.082 | 0.068 |
| | (0.025) | (0.025) | (0.023) | (0.023) | (0.055) | (0.052) |
| T3: popularity (based on 2018) | 0.056** | 0.059** | 0.070*** | 0.070*** | 0.079 | 0.081* |
| | (0.027) | (0.027) | (0.025) | (0.024) | (0.049) | (0.047) |
| Observations | 6,861 | 6,861 | 6,861 | 6,861 | 6,861 | 6,861 |
| R-squared | 0.016 | 0.047 | 0.064 | 0.092 | 0.010 | 0.044 |
| Controls | No | ✓ | No | ✓ | No | ✓ |
| Lagged Dep. Var. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $H_0$: Pooled T2/T3 = T1 (p-value) | .316 | .242 | .053 | .04 | .914 | .964 |
| $H_0$: T1=T2 (p-value) | .261 | .21 | .061 | .078 | .952 | .936 |
| $H_0$: T1=T3 (p-value) | .543 | .453 | .147 | .078 | .903 | .869 |
| Mean of Dep. Variable in Control | .174 | .174 | .437 | .437 | 1.51 | 1.51 |
| $H_0$: T2=T3 (p-value) | .652 | .668 | .820 | .853 | .962 | .824 |

- Exposure to the intervention: diffusion models directly applicable
- Counseling impacts larger for T2 and T3

# Main behavioral outcomes

| | Pregnancy in last 2 years | | Had sex and used contr. last time | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| VARIABLES | | | | |
| T1: school selects (based on 2018) | -0.005 | -0.003 | 0.007 | 0.013 |
| | (0.008) | (0.008) | (0.014) | (0.012) |
| T2: network centrality (based on 2018) | -0.020** | -0.018** | 0.028* | 0.034** |
| | (0.008) | (0.008) | (0.015) | (0.014) |
| T3: popularity (based on 2018) | -0.013 | -0.010 | 0.018 | 0.018 |
| | (0.008) | (0.008) | (0.017) | (0.017) |
| Observations | 6,861 | 6,861 | 6,861 | 6,861 |
| R-squared | 0.052 | 0.088 | 0.127 | 0.162 |
| Controls | No | ✓ | No | ✓ |
| Lagged Dep. Var. | ✓ | ✓ | ✓ | ✓ |
| $H_0$: Pooled T2/T3 = T1 (p-value) | .099 | .097 | .299 | .358 |
| $H_0$: T1=T2 (p-value) | .064 | .064 | .237 | .192 |
| $H_0$: T1=T3 (p-value) | .304 | .339 | .563 | .785 |
| Mean of Dep. Variable in Control | .072 | .072 | .567 | .567 |
| $H_0$: T2=T3 (p-value) | .463 | .377 | .606 | .418 |

- Converting knowledge to action function of information and norms
- Together, network-based targeting more effective than T1 at reducing pregnancy

## Take-aways

Network-based selection more effective than status quo benchmark

- Cannot distinguish selection on popularity vs. centrality

Use a network model to further tease out mobilizer effectiveness

- Centrality useful for spreading info (as in prior literature)
- Popularity required for norm change $\implies$ behavior change

While costly network elicitation used in the experiment, useful for model, potential for "shortcuts" in scale-up

- Popularity easy to measure
- Can ask "gossip centrality" questions as in Banerjee et al (2018)

# Rational Aggregation

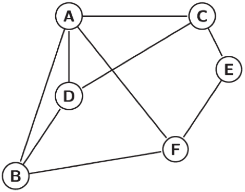
Suppose all nodes $i$ in a network receive some iid signal $p_i$ at $t = 0$

- Agents can arrive at the correct beliefs if they come to learn entire vector $p$
- Set of models that consider learning on networks where agents communicate information tagged with its source: *tagging*
  - See., e.g., Acemoglu et al (2014), Mobius et al (2015)
  - Here Baysian learning is the right benchmark, communicate elements of $p$ by diffusion, aggregate with Bayes rule.

# Rational Aggregation Without Tagging

Suppose agents can't tag information source (constraints on dimensionality of what can be passed)

- Need to infer the meaning of signals from neighbors on the network




- F talks to E, B, A
- When weighting signals, rational Bayesian F needs to figure out independent component of each node's information vs. common component from upstream node's info
  - e.g., D's signal will be reflected in their messages
- Requires complete knowledge of network structure

# Aside: Aggregating One's Own Signals

Even before jumping to signal aggregation on a network, Bayesian learning has strong predictions for how people learn from their *own* signals.

- Individuals optimally aggregate all information they experience before making decisions
- Importantly they need to attend to each dimension of data they collect

"Learning through Noticing" (2014) Hanna, Mullainathan, Schwartzstein

- Context: Seaweed farming in Indonesia
- Cultivated by taking raw seaweed and cutting it into pods, which are then planted at intervals along the ocean floor.
- Size of the pods and distance between them are important choices.
- Short crop cycle of 35 to 40 days, ample opportunity to learn through experimentation.

# Aside: Aggregating One's Own Signals

Experiment:

- Enroll farmers in trials to experiment with production. (e.g., change planting techniques systematically over different input dimensions). Supervised by researchers.
- Finding - farmers were using the wrong pod size!

# Aside: Aggregating One's Own Signals

Experiment:

- Enroll farmers in trials to experiment with production. (e.g., change planting techniques systematically over different input dimensions). Supervised by researchers.
- Finding - farmers were using the wrong pod size!

What did farmers do with the results?

- Experimentation alone did not change practices
- But - providing farmers with easy-to-digest trial summary did lead to behavior change

Not very promising for Bayesian learning models

- Justification for behavioral or naive learning models

# DeGroot Model

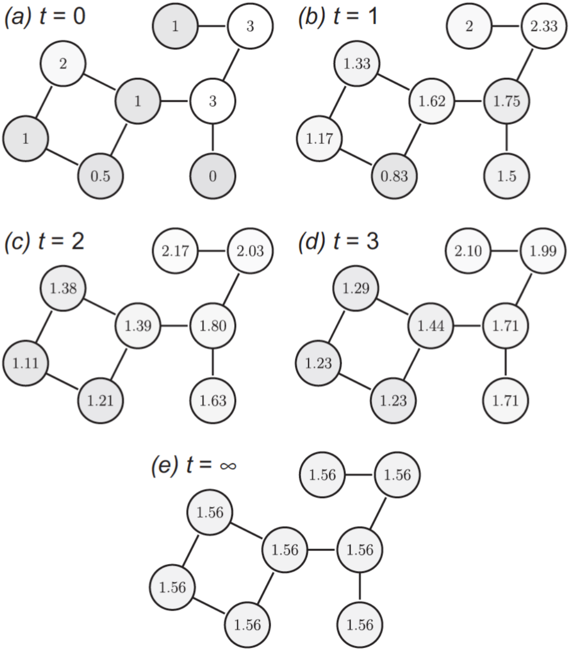
Workhorse model of naive learning: DeGroot

- $n$ nodes interact on network $T$
- $T$ stochastic, meaning all rows sum to 1
  - e.g., $T_{ij} = \frac{A_{ij}}{d_i}$, $A_{ij} \in \{0, 1\}$
- Behavioral updating rule, time $t$ beliefs: $p^{(t)} = Tp^{(t-1)}$
  - So, $p^{(t)} = T^t p^{(0)}$
  - Belief is average of beliefs of network connections

Theoretical results (Golub and Jackson 2010) under regularity conditions

- Society converges to the same limit belief
- That limit belief converges to the truth so long as no nodes have outsized influence
- So DeGroot "works well" in the limit.

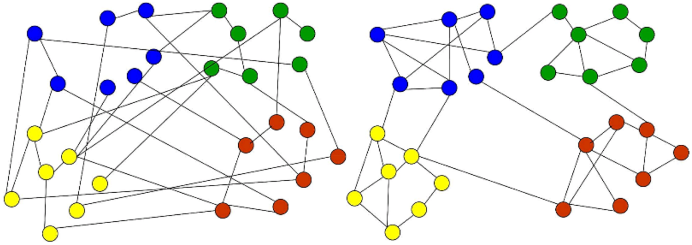
**Figure 3:** DeGroot updating in a sample social network



# Homophily and Consensus Time

Golub and Jackson (2012) consider how convergence time is a function of network structure

- Networks that exhibit *homophily*, with inward-looking groups can be very slow to converge to consensus



(a) (b)

Figure 1: Islands networks with low and high homophily are shown in (a) and (b), respectively.

- Prediction: learning slower in network b) vs. a)

# Application (in progress)

Arun Chandrasekhar, M.R. Sharan and I are using this concept in the context of caste-based reservation in India

- What is the effect of political representation for historically disadvantaged groups (scheduled castes: SCs) on social structure?
- Context - local rural governments (Gram Panchayats: GPs) in Bihar
- RD-based empirical strategy based on assignment algorithm
- Network surveys

# Within- and Cross-group Linking

| | Link Rate to non-SCs | | |
|---|---|---|---|
| | All | SC | non-SC |
| SC Reservation | -0.298 | -1.969 | -0.041 |
| | (0.066) | (0.015) | (0.034) |
| | (-0.615, 0.020) | (-2.220, -1.718) | (-0.316, 0.235) |

| | Link Rate to SCs | | |
|---|---|---|---|
| | All | SC | non-SC |
| SC Reservation | 0.476 | 2.874 | -0.216 |
| | (0.090) | (0.649) | (0.000) |
| | (0.040, 0.913) | (1.487, 4.261) | (-0.441, 0.009) |

- SC to non-SC $\downarrow\downarrow$ ; SC to SC $\uparrow$ 64%
- non-SC to non-SC no change; non-SC to SC $\downarrow$ 45%
- Consistent with increase in homophily?

# Reservation Increases Homophily

- How do $\lambda_2$ (related measure of homophily) and $CT$ respond to reservation?
- Network, $H := H(P)$ with $P_{5\times 5}$ entries $P_{kk'}$ cross-subcaste link rates

| | **homophily** | |
|---|---|---|
| | $\lambda_2$ | log Consensus Time |
| SC Reservation | 0.355<br>(0.130, 0.580) | 2.165<br>(1.667, 2.664) |
| Control.mean | 0.718 | 2.841 |

- Homophily ↑; time to convergence takes 9× longer
- Does actual learning look worse in reserved constituencies?
  - Under "seeding", info should have a harder time crossing caste boundaries

## Social Learning Friction - Policy Knowledge

- During the pandemic, ASHAs repurposed from TB / infant health to COVID.
    - But many did not know who the ASHA even was...

# Social Learning Friction - Policy Knowledge

- During the pandemic, ASHAs repurposed from TB / infant health to COVID.
  - But many did not know who the ASHA even was...

| | **Know ASHA Worker** |
|---|---|
| | ALL |
| SC Reservation | -0.486 |
| | (-0.711, -0.261) |

- Info about scholarships seeded with teachers. Lower diffusion?

| | **Child Received Scholarship From the Gov't in last 2 years?** | | |
|---|---|---|---|
| | All | SC | non-SC |
| SC Reservation | -0.212 | -0.590 | -0.087 |
| | (-0.437, 0.013) | (-0.815, -0.365) | (-0.312, 0.138) |
| Control.mean | 0.462 | 0.417 | 0.5 |

# Wrap-Up: Diffusion and Aggregation

- Diffusion essential process for information flow
- Relevant for many settings: agricultural extension, spreading info about new financial products, new government programs, job opportunities etc.
- Large gains empirically from targeting well
    - Empirically, notions of centrality linked to number of times people hear a piece of information work well
    - Has a complex contagion flavor, already
- Complex contagion on take-up seems like the appropriate model for some types of risky investment decisions
- Aggregation very difficult, behavioral models likely more appropriate
- Even with wisdom, network features affect quality of learning

# Roadmap

1. Value of Networks
2. Introduction to Networks
3. Information Diffusion and Aggregation
4. Network - Market Interactions

# Formal Finance when Informal Finance is Already There

Vibrant informal market for loans in developing countries:

- Moneylenders
- Family and risk sharing network
- Trade credit

How do new sources of formal credit interact with existing informal sources and social relationships?

- Is microfinance improving financial inclusion? Are people gaining access to credit who would otherwise be unbanked?
- OR, is microfinance simply lowering the cost of credit (interest rate) without expanding overall credit access?
- Is microfinance crowding out or crowding in network relationships?

Important question because financial inclusion policy often enacted through preferential lending and subsidies

# Banerjee, Breza, Chandrasekhar, Duflo, Kinnan and Jackson (2023)

We combine data from two "experiments"

- "Diffusion of Microfinance" natural experiment (Banerjee et al 2014):
  - Baseline network survey (13 dimensions of relationships) collected in 75 villages
  - Some villages added microfinance (post-network survey)
  - 43 out of 75 (not random)
  - Wave 2 Network survey collected 5-6 years later
- Hyderabad MF RCT

Research Question: How does network change because of microfinance? Are there GE impacts, even for those who aren't interested/eligible for MF?

# Link-Level Analysis

- Identify which households would tend to have gotten loans in non-MF villages/neighborhoods
- Use baseline predictors of access to microfinance in a random forest model
- Allows comparison of likely loan takers/non takers across MF and non-MF areas
- Two types of households: $H$ and $L$
- how does microfinance exposure affect the formation of links across types ($H$ and $L$) of households?
    - $LL$, $LH$, $HH$ denote link by type pairs

# Link-Level Analysis: Karnataka

| | (1) Linked Post-MF | (2) Linked Post-MF | (3) Linked Post-MF | (4) Linked Post-MF |
|---|---|---|---|---|
| Microfinance | −0.058 | −0.060 | −0.023 | −0.021 |
| | (0.018) | (0.020) | (0.008) | (0.008) |
| | [0.002] | [0.003] | [0.006] | [0.007] |
| Microfinance × *LH* | 0.009 | 0.001 | 0.007 | 0.007 |
| | (0.015) | (0.014) | (0.004) | (0.004) |
| | [0.573] | [0.936] | [0.120] | [0.081] |
| Microfinance × *HH* | 0.039 | 0.023 | 0.009 | 0.013 |
| | (0.022) | (0.021) | (0.007) | (0.006) |
| | [0.086] | [0.280] | [0.206] | [0.040] |
| Observations | 57,376 | 57,376 | 846,561 | 846,561 |
| Linked Pre-MF | Yes | Yes | No | No |
| Controls | | ✓ | | ✓ |
| Depvar Mean | 0.441 | 0.441 | 0.0636 | 0.0636 |
| *LL*, Non-MF Mean | 0.482 | 0.482 | 0.0753 | 0.0753 |
| MF + MF x *LH*= 0 p-val | 0.014 | 0.009 | 0.015 | 0.015 |
| MF + MF x *HH*= 0 p-val | 0.361 | 0.09 | 0.101 | 0.233 |
| MF + *LH*xMF = MF + *HH*xMF p-val | 0.137 | 0.275 | 0.641 | 0.231 |

- Links fall for *LL* pairs, actually a stronger decline than *LH* or *HH* pairs

# Triads of Nodes: Karnataka

What about triples? Maybe *LL*s that are dropping are linked to an *H* (*LLH* triads)

| | (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 | −0.069 | −0.085 | −0.081 |
| | (0.029) | (0.026) | (0.023) | (0.019) |
| | [0.008] | [0.008] | [0.000] | [0.000] |
| Microfinance × *LLH* | 0.026 | 0.014 | 0.043 | 0.034 |
| | (0.021) | (0.019) | (0.018) | (0.015) |
| | [0.228] | [0.463] | [0.015] | [0.024] |
| Microfinance × *LHH* | 0.054 | 0.026 | 0.057 | 0.039 |
| | (0.030) | (0.024) | (0.025) | (0.018) |
| | [0.072] | [0.274] | [0.022] | [0.029] |
| Microfinance × *HHH* | 0.093 | 0.045 | 0.087 | 0.058 |
| | (0.042) | (0.036) | (0.031) | (0.026) |
| | [0.028] | [0.206] | [0.006] | [0.023] |
| 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 |

- Even the *LLL* triples fall!
- Consistent with microfinance imposing a *global* externality on network formation
- Propose a model where individuals must pay an effort cost to

# Link-Level Analysis: Hyderabad

| | (1) Prob. Linked | (2) Prob. Linked |
|---|---|---|
| Microfinance | $-0.006$ | $-0.006$ |
| | (0.003) | (0.003) |
| | [0.023] | [0.035] |
| Microfinance $\times$ $HH$ | $-0.009$ | $-0.009$ |
| | (0.009) | (0.008) |
| | [0.296] | [0.269] |
| Microfinance $\times$ $LH$ | 0.003 | 0.002 |
| | (0.003) | (0.003) |
| | [0.432] | [0.470] |
| Observations | 141,990 | 141,990 |
| Controls | No | Yes |
| Depvar Mean | 0.0255 | 0.0255 |
| LL, Non MF Mean | 0.0268 | 0.0268 |
| MF + MF $\times$ $HH = 0$ p-val | 0.097 | 0.081 |
| MF + MF $\times$ $LH = 0$ p-val | 0.458 | 0.396 |
| MF + MF $\times$ $HH =$ MF + MF x $LH$ p-val | 0.049 | 0.047 |

- Similar patterns: LL households lose links because of microfinance

## Measuring Insurance Value

Recall "Townsend Regression" (Townsend, 1994)

$$c_{ivt} = \alpha + \beta y_{ivt} + \mu_{vt} + \epsilon_{ivt}$$

- Under full insurance $\beta = 0$.
- More generally $\mathrm{corr}(c_i, y_i | C_v) = 0$.

Treatment interactions

$$\begin{aligned} c_{ivt} = {} & \alpha + \beta_1 y_{ivt} + \beta_2 y_{ivt} \times \mathrm{Treatment}_v \\ & + \beta_3 H_i \times y_{ivt} + \beta_4 y_{ivt} \times H_i \times \mathrm{Treatment}_v \\ & + \tau H_i \times \mathrm{Treatment} + \gamma H_i + \delta \mathrm{Treatment}_v + \mu_{vt} + \epsilon_{ivt} \end{aligned}$$

- $\beta_2 > 0$: *increase* in income-consumption correlation for $L$s when network gets credit access

# $L$s lose consumption smoothing

| | (1) Expend.: Total | (2) Expend.: Non-Food |
|---|---|---|
| Household Income per capita | 0.111 | 0.059 |
| | (0.027) | (0.021) |
| | [0.000] | [0.005] |
| Microfinance $\times$ Income | 0.069 | 0.080 |
| | (0.041) | (0.034) |
| | [0.098] | [0.018] |
| Household Income per capita $\times H$ | 0.072 | 0.032 |
| | (0.051) | (0.034) |
| | [0.157] | [0.351] |
| Microfinance $\times$ Income $\times H$ | -0.121 | -0.107 |
| | (0.074) | (0.060) |
| | [0.103] | [0.075] |
| Observations | 10452 | 10361 |
| Test: MF $\times$ Inc + MF $\times$ Inc $\times H$ | 0.348 | 0.546 |

- Goal: If $L$s lose links, do they lose insurance?
  - Is $c_i$ more correlated with $y_i$ with MF?
  - Use Hyderabad endline consumption, income data
- Townsend 1994-type reg of consumption on:
  - own income
  - treatment
  - $H$ type (w/ interactions)
- Finding:
  - $L$s experience a relative increase in corr$(c_i, y_i)$
  - $H$s experience no change
  - $L$ income unaffected by MF (unreported)