



PROGRAMME EVALUATION AND SOCIAL NETWORKS

ESRC Research Methods Festival, July 2012

PEPA Programme Evaluation
for Policy Analysis

PROGRAMME EVALUATION FOR POLICY ANALYSIS: OVERVIEW

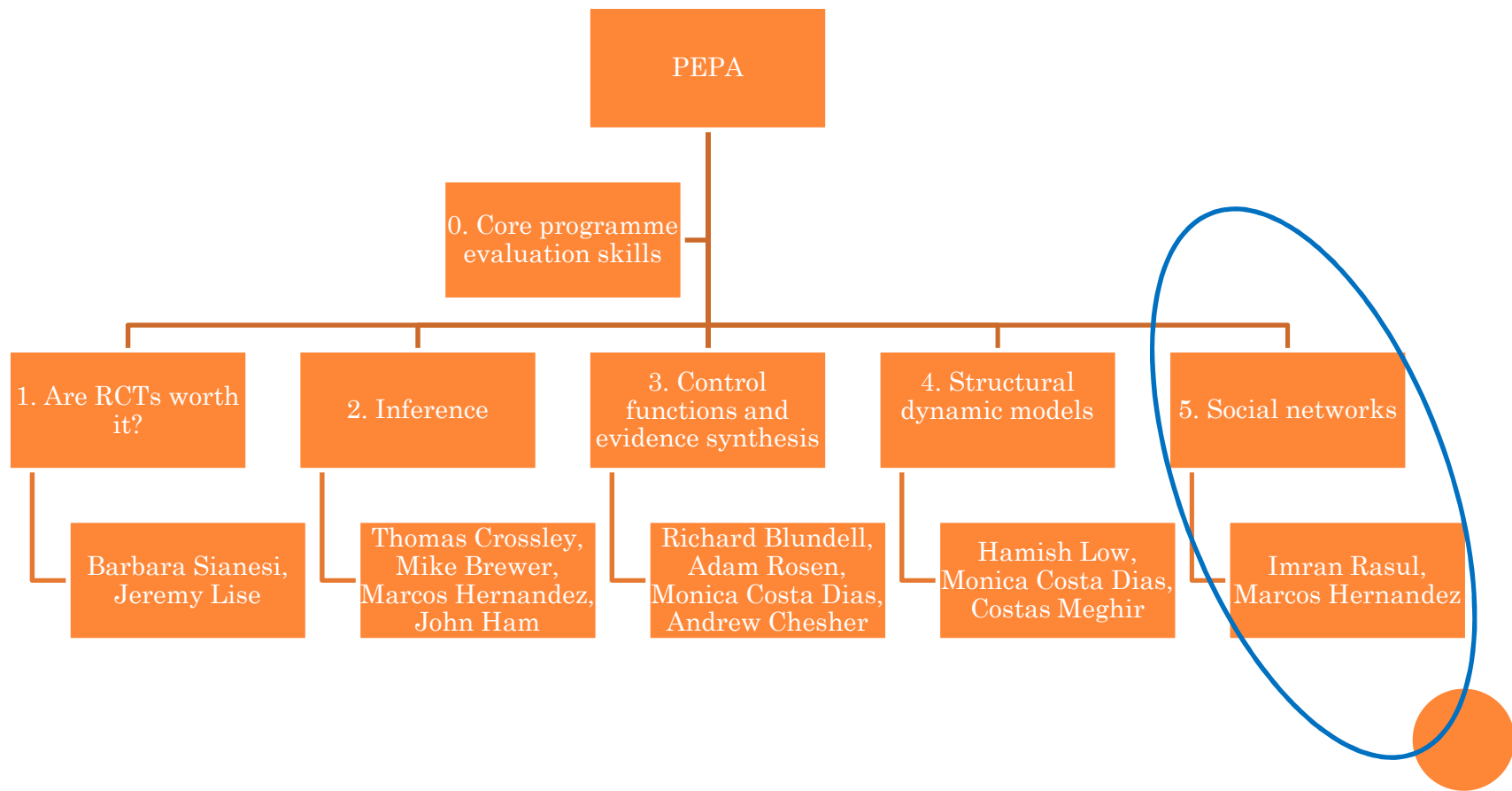
- PEPA is about ways to do, and ways to learn the most from, “programme evaluation”

“government policies”
(although can often generalise)

“estimating the casual impact of”



PEPA: OVERVIEW



OVERVIEW

- Networks in economics
- Networks and programme evaluation
- Three case studies:
 - Mexico
 - Malawi
 - Bangladesh

- Acknowledgements: slides based on work by Daron Acemoglu, Asu Ozdaglar and Sanjeev Goyal

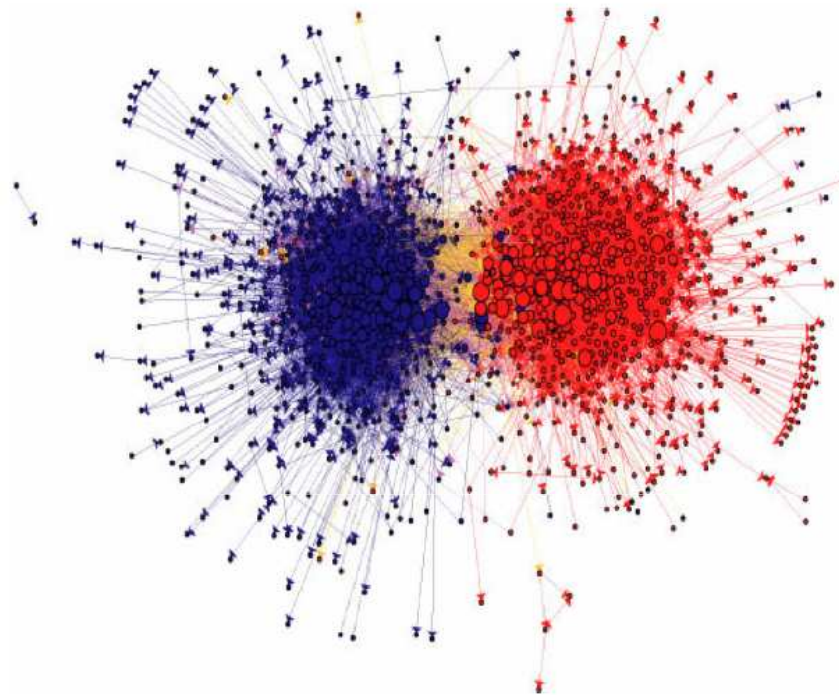


NETWORKS

- Networks represent the **interaction** structure between units (nodes)
- In economic networks, these nodes can be individuals, firms, governments for example
- Study of networks useful for understanding many types of interaction:
 - Information transmission
 - Friendship/trust
 - Diffusion of ideas
 - Trade and exchange
 - **Response to programme interventions**
 - Spillovers to non-beneficiaries, distributional impacts



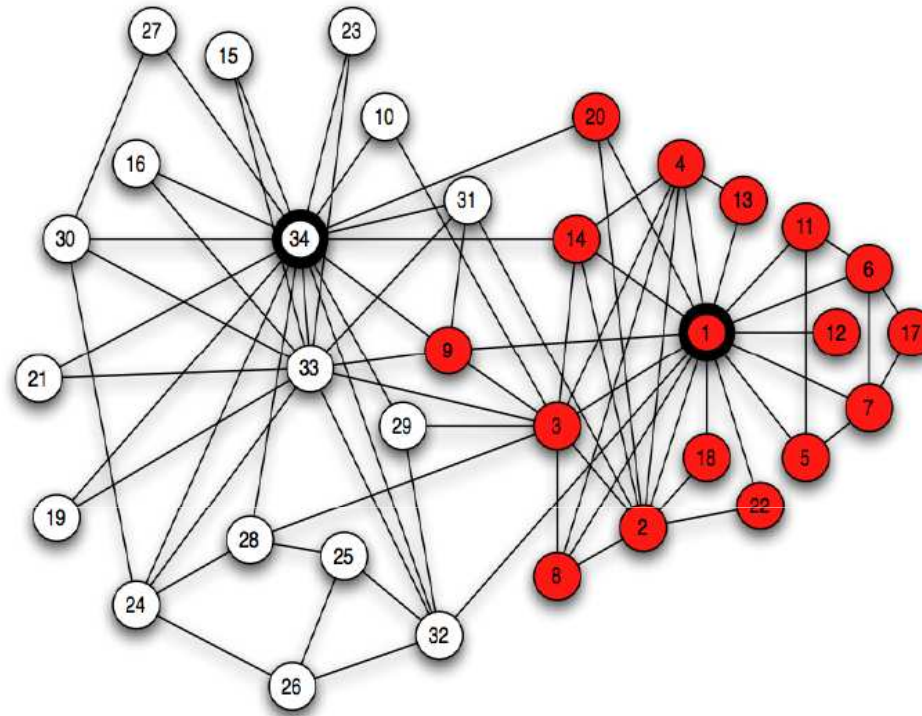
EXAMPLE 1



Information transmission: Networks structure of political blogs prior to the 2004 US Presidential Election: two separated clusters
[Adamic and Glance 2005]

- Programme evaluation: those in information networks of the direct beneficiaries might also be affected

EXAMPLE 2

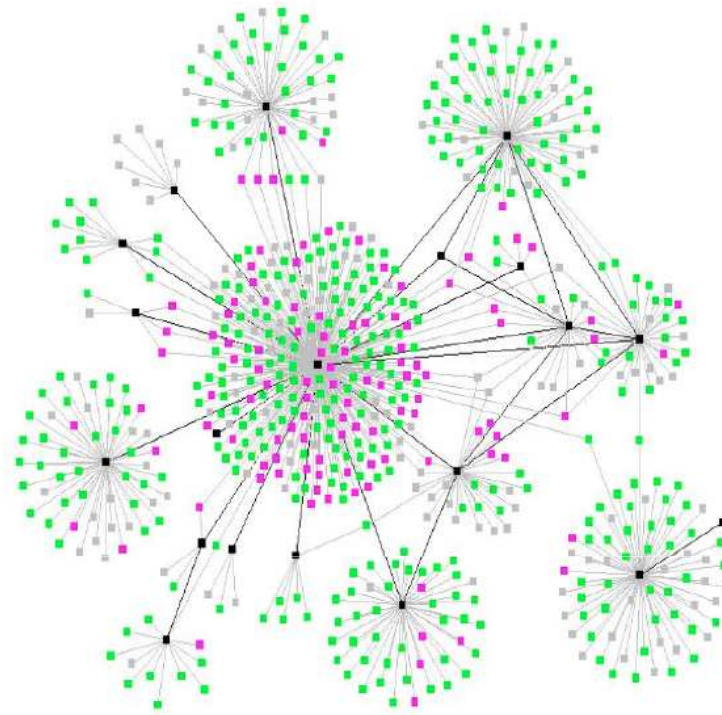


Social network of **friendships** in a 34-person karate club: club eventually split [Zachary 1977]

Programme evaluation: is there an **optimal** node to intervene on, to maximize social impact?



EXAMPLE 3

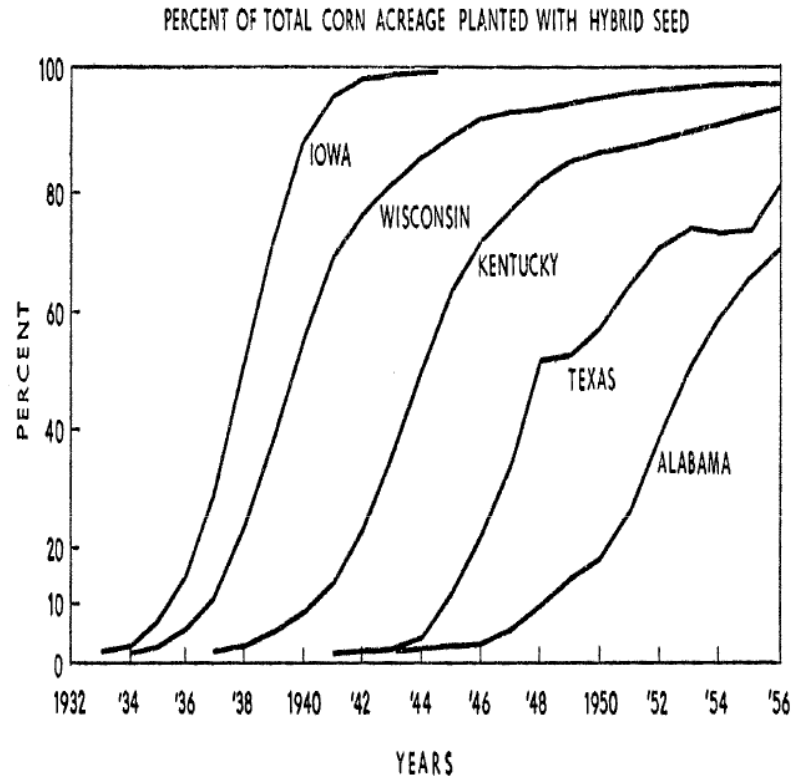


The spread of an **epidemic** disease (TB shown) [Andre et al. 2007]:

- key distinction between economic and biological models is that in economic models agents in part driven by **strategic/choice** considerations
- leads to the use of **game theory** to analyze behavior within networks
- Programme evaluation: if this pattern can be foreseen can policy react?



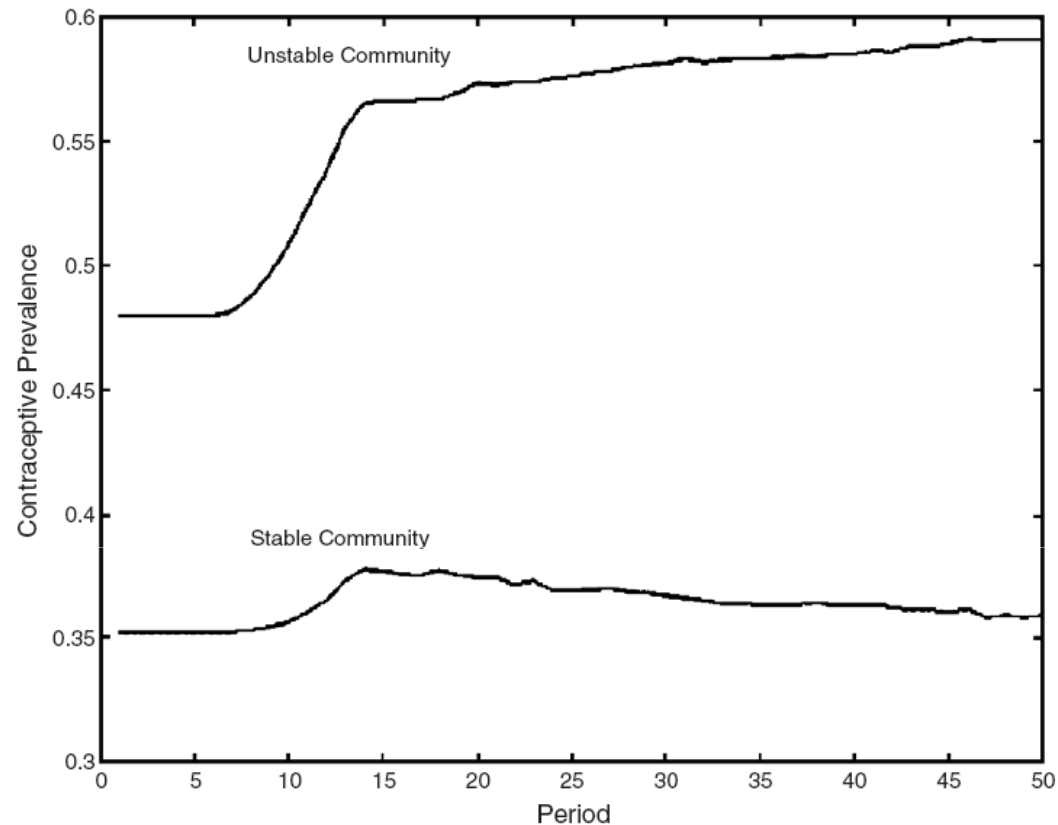
EXAMPLE 4A



Technology adoption: percentage of total corn acreage planted with hybrid seed [USDA]
Motivated Griliches [1957, 1958] seminal studies



EXAMPLE 4B

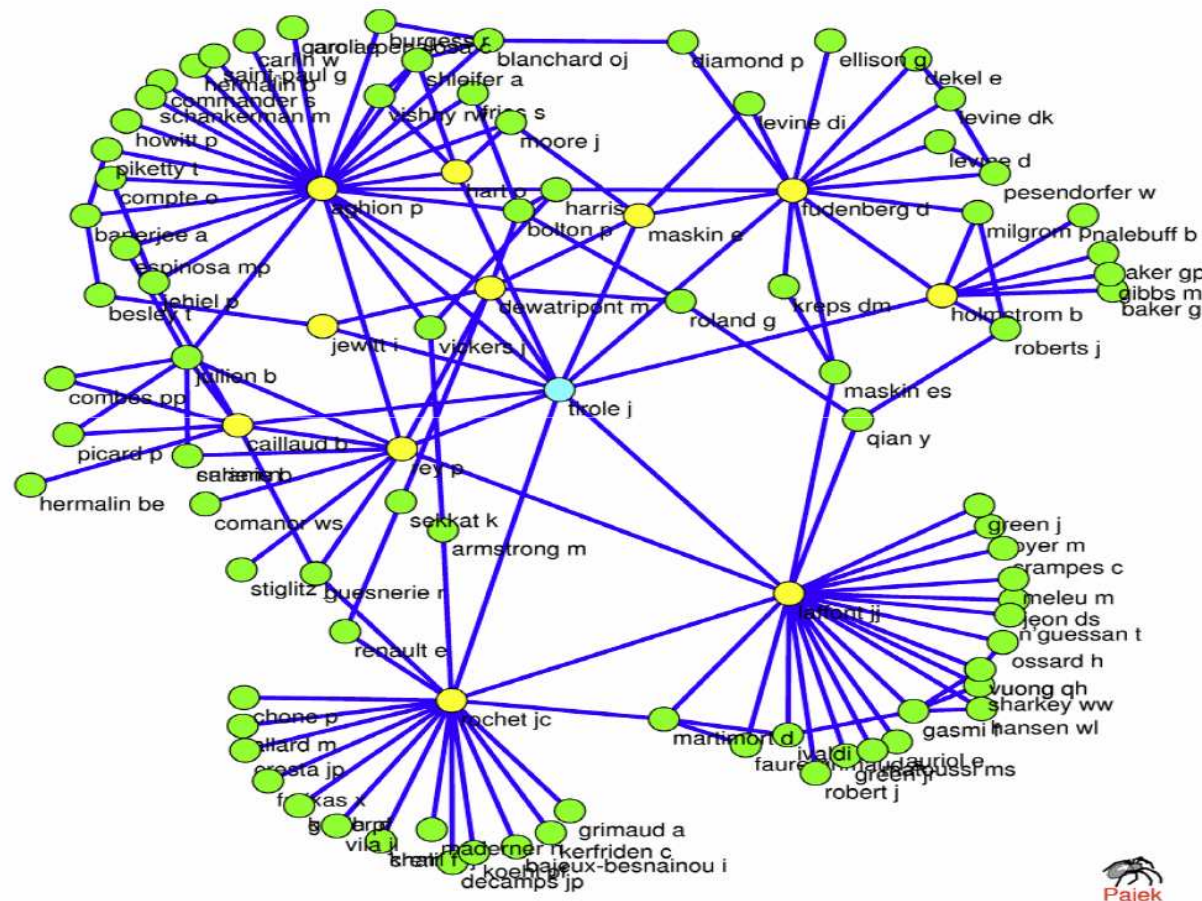


Technology adoption: use of modern contraceptive methods in rural Bangladesh [Munshi and Myaux [2006]

- some communities reach tipping points and switch behavioural norms
- Programme evaluation: why the same policy might not be equally effective everywhere



EXAMPLE 5



Co-authorship network of Jean Tirole in 1990s

- Programme evaluation: links not formed at random – can be **endogenous**, might also respond to interventions [last example]



SMALL WORLD EXPERIMENT

- Sociologist Stanley Milgrom originally studied the “Small World Problem” in 1967
- Asked certain residents of Wichita and Omaha to contact and send a folder to a target person by sending it to an acquaintance, who would do the same etc. until the target was reached
- **Research question:** how many links (intermediate nodes) would be required to reach the target?
- **Result:** 42 out of 160 letters made it to the target, with a median number of intermediate nodes equal to 5.5



INTERPRETING SMALL WORLDS

- Suppose each node has λ neighbours
- Each neighbour will have λ neighbours
- Suppose **unrealistically** that my neighbours don't have neighbours in common
- Hence in two steps you can reach $\lambda \times \lambda$ neighbours
- In d -steps can reach λ^d other nodes
- Suppose network has $n = \lambda^d$ nodes, so average distance is,

$$d = \frac{\ln n}{\ln \lambda}$$

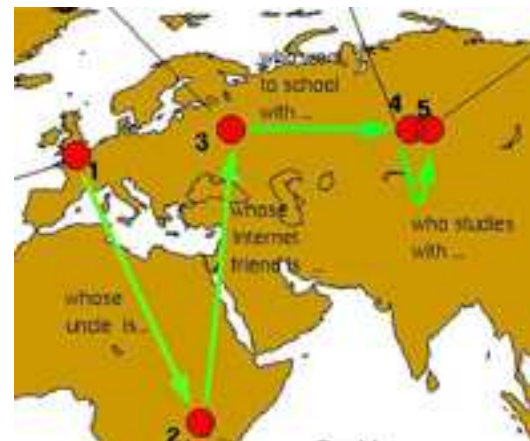


GLOBAL DEGREES OF SEPARATION

- Commonly held belief of “six degrees of separation” between any two individuals on the planet,

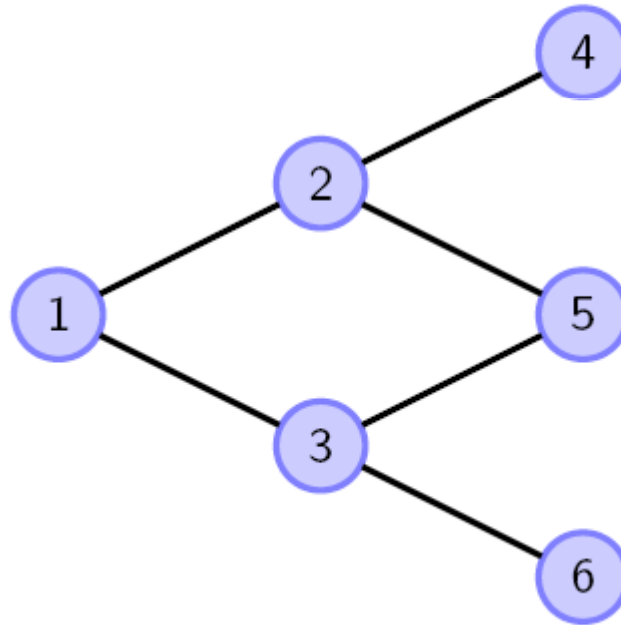
$$d = \frac{\ln n}{\ln \lambda} = \frac{\ln 7000000000}{\ln \lambda}$$

- So if $d=6$, implies λ is around 19
- Experiment recently repeated by Duncan Watts, sociologist at Columbia, using **email** technologies



INTERPRETING SMALL WORLDS

- But this method rules out triadic relations and clustering phenomena, that we have seen are common in some of the visual examples



NETWORKS IN ECONOMICS AND SOCIOLOGY

- Focus in sociology on group interactions so network structure is important
 - notions of social capital, power and leadership
- Economics about allocation of scarce resources
 - trade, cooperation, competition, information exchange, technology adoption etc.
- Neoclassical economics studies one of two extremes:
 - markets: all interactions feasible and anonymous, e.g. GE theory
 - games among few players: predetermined player identities, e.g. auctions



NETWORKS IN ECONOMICS

- Social structures viewed of as being important in **developing** country contexts
 - replace missing or imperfect markets
 - in 'modern' economies, view was that trade takes place among anonymous agents meeting in markets
 - examples in PEPA are from LDCs, but could expand in the future
- With greater recognition of **information asymmetries**, role of social networks to explain behavior in modern economics
 - informal institutions also matter in developed economies
- Social networks will interplay with programme interventions in rich and poor economics



NETWORKS AS GRAPHS 1

- Can mathematically represent networks with graphs, that formalize the pattern of links between nodes
- Graphs can be directed or undirected
- Links can be weighted or unweighted, depending on whether links differ in importance
- A directed (unweighted graph) is,

$$G=(N,E)$$

- N = set of nodes
- E = set of edges



NETWORKS AS GRAPHS 2

- $j \in N$ if j is a node in this network
- $(i,j) \in E$ if there is a link from i to j
- In a directed graph, this does **not** imply $(j,i) \in E$
- Can also use notion $g_{ij}=1$ if $(i,j) \in E$ and $g_{ij}=0$ otherwise
- In a weighted graph, $g_{ij}>0$ would measure the strength of the link from i to j
- Can then use these building blocks to capture **characteristics** of nodes in a network
- In turn, these characteristics might determine how nodes are affected by policy interventions



POWER IN A NETWORK

- A measure of power that takes into account the location of nodes within the network is the “**betweenness**” measure
- $P(i,j)$ = number of shortest paths connecting i to j
- $P_k(i,j)$ = number of shortest paths between i and j that **include k** ,

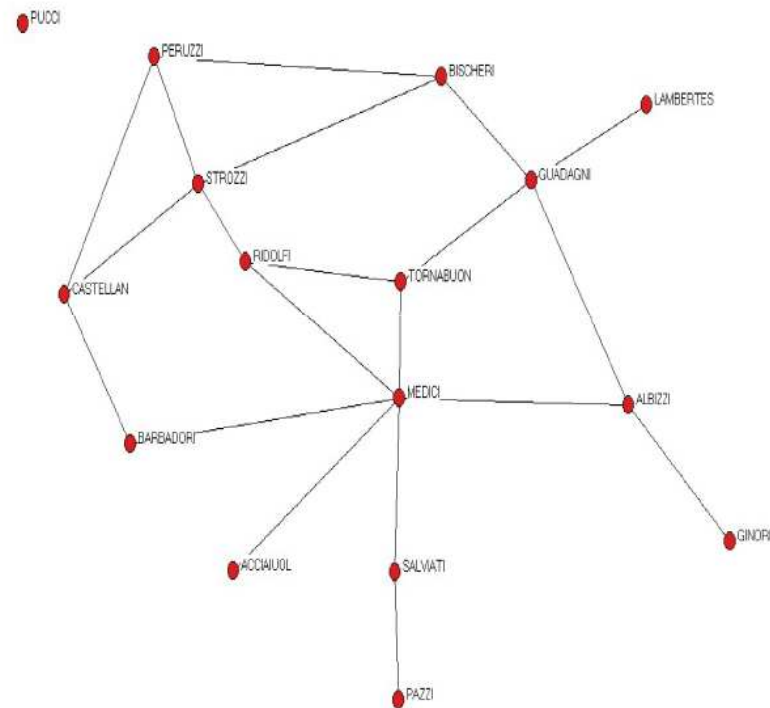
$$B_k = \sum_{(i,j) \in E; i \neq j, k \notin \{i,j\}} \frac{P_k(i,j) / P(i,j)}{(n-1)(n-2) / 2}$$

- Convention is that $P_k(i,j)/P(i,j)=0$ if $P(i,j)=0$
- Betweenness measure gives, for each pair of nodes, the fraction of shortest paths that go **through node k**



EXAMPLE OF POWER IN A NETWORK

- Power of families in 15th century Florence [Padgett and Ansell 1993]
- How did Medici's become so influential in politics and economics?
- Betweenness measure for Medicis is .522
- Next highest family is .255
- Will later see village figures for extended family networks in rural Mexico



OTHER IMPORTANT ELEMENTS OF NETWORK STRUCTURE

- **Degree distribution:** average degree is very small relative to the number of nodes, and huge inequality
- **Clustering:** typically high in social networks; if links formed at random then in large n -node network with average degree k , clustering would be roughly k/n
 - in economics co-authorship network (example 6), clustering coefficient is .157 (7000 times larger than would be expected with random matching)
- **Centrality:** Bonacich centrality is key measure explaining behavior in some models; also used to identify **key players** to remove from networks for maximal policy effect [Ballester et al. 2006]



RESEARCH QUESTION 1A: LINK FORMATION

- What is the process by which links form and are broken?
- Key features of process of **link formation**:
 - Linking is a decision
 - Externality/spillovers: link between i and j affects the payoff of k and the payoffs to k from linking to i or j
 - Combine: games of network formation
- Key **modelling** issues:
 - Payoffs: linking generates rewards and entails costs
 - Power: who decides on the link (uni or bi-directional)
 - Information: what do I know about other players and the network when I form a link?



RESEARCH QUESTION 1B: LINK FORMATION

- Are more links always better?
- Costs:
 - trade-off with trust, investments per link, ability to punish non-cooperators in a network
- Benefits:
 - provides access to new information, additional resources
 - ability to change social norms, shift from low to high equilibrium payoffs
- Policy 1: whether and how process of link formation responds to policy interventions?
- Policy 2: are all socially optimal links formed?



RESEARCH QUESTION 2: DIVERSITY

- **‘Wisdom of Crowds’**: combining information of many leads to better decisions
 - especially so if crowd has diverse experiences and perspectives [Galton, de Condorcet]
 - suggests large networks can reach more accurate decisions
- Concern is that ‘groupthink’ is also more prevalent in large groups (a form of herding)
- Cooperation (free-riding) and coordination harder to achieve in larger groups
- **Arrows Impossibility Theorem**: impossible for a group to have decision rule that is efficient and non-dictatorial and that satisfies the IIA
- Policy: do interventions have heterogeneous impacts across networks?



RESEARCH QUESTION 3: COMMUNICATION

- Have advances in information and communication technology changed the nature of social networks?
 - Columbia small worlds experiment
 - political blogs example: no guarantee that more diverse information is acquired
 - can greater access to information increase ‘herding’?, i.e. excessive copying of others’ behavior
- Policy: how do policy impacts vary with communication technologies?



RESEARCH QUESTION 4: EMPIRICS

- PEPA focus is on how the impact of policy interventions interacts with the social networks of communities subject to the intervention
- Important implications for:
 - evaluating policy impacts
 - equilibrium effects of policies where the policies might impact non-eligibles through network structures
 - understanding the distributional consequences of policy
 - understanding why the same policy might have different impacts across locations



EMPIRICAL EXAMPLES

- Progresa and extended family networks in Mexico
- The Women's Group Programme in Malawi and networks of brothers and sisters
- The Ultra-Poor Programme in Bangladesh and village networks



De Giorgi *et al* (2010) Family Networks and School Enrolment: Evidence From a Randomized Social Experiment

- descriptive evidence on the presence and characteristics of a household's extended family within close geographic proximity
- econometric evidence to identify the causal effect of the extended family on the response to a CCT program, Progresas, on the household's schooling choices
 - basic idea is to use *Progresas* as an exogenous shifter of the net resources available to the household **and** to the family network
 - families are able to enforce implicit contracts of resource sharing
 - hence program can induce differential responses between households embedded within family networks and those that are socially isolated



RESEARCH QUESTIONS

- document the proportions of households that are embedded within a family network, and that are isolated
- do husbands and wives differ in the extent to which members of their extended family live in close proximity?
- do the characteristics of extended family members influence the response to *Progresa* in terms of the secondary school enrolment of children in the household?
 - compare responses of eligible households embedded in family networks vis-à-vis responses of eligible but isolated households
 - for connected households, whether and why the characteristics of specific intra-generational (sibling) and inter-generational (parent-child) links matter for responses to *Progresa*



THE PROGRESA PROGRAM

- *Progresa* social assistance program provides cash grants to women in the household
- paid bimonthly, one component is conditional on children attending classes at least 85% of the previous 60 days
- these transfers correspond approximately to **one half to two thirds** of the full time child wage [Schultz 2004]
 - the most constrained households should not respond to conditional transfers related to **secondary** enrolment
 - cash transfers for **primary** enrolment act more as an **unconditional** transfer [pure income effect]
 - health and nutrition transfer components somewhere in between [those with no children may be eligible]



THE EVALUATION DATA

- household panel data collection in 506 geographically remote villages every six months, Oct 97 (wave 1) to Nov 99 (wave 5)
- randomization is conducted at the village level: 186 out of 506 villages randomized out
- 52% of households classified as eligible (poor) based on their poverty status in October 1997
- *Progresa* transfers first distributed in May 1998
- within each village eligibles and non-eligibles are surveyed [complete census]
- focus on couple headed households [85% of all households]

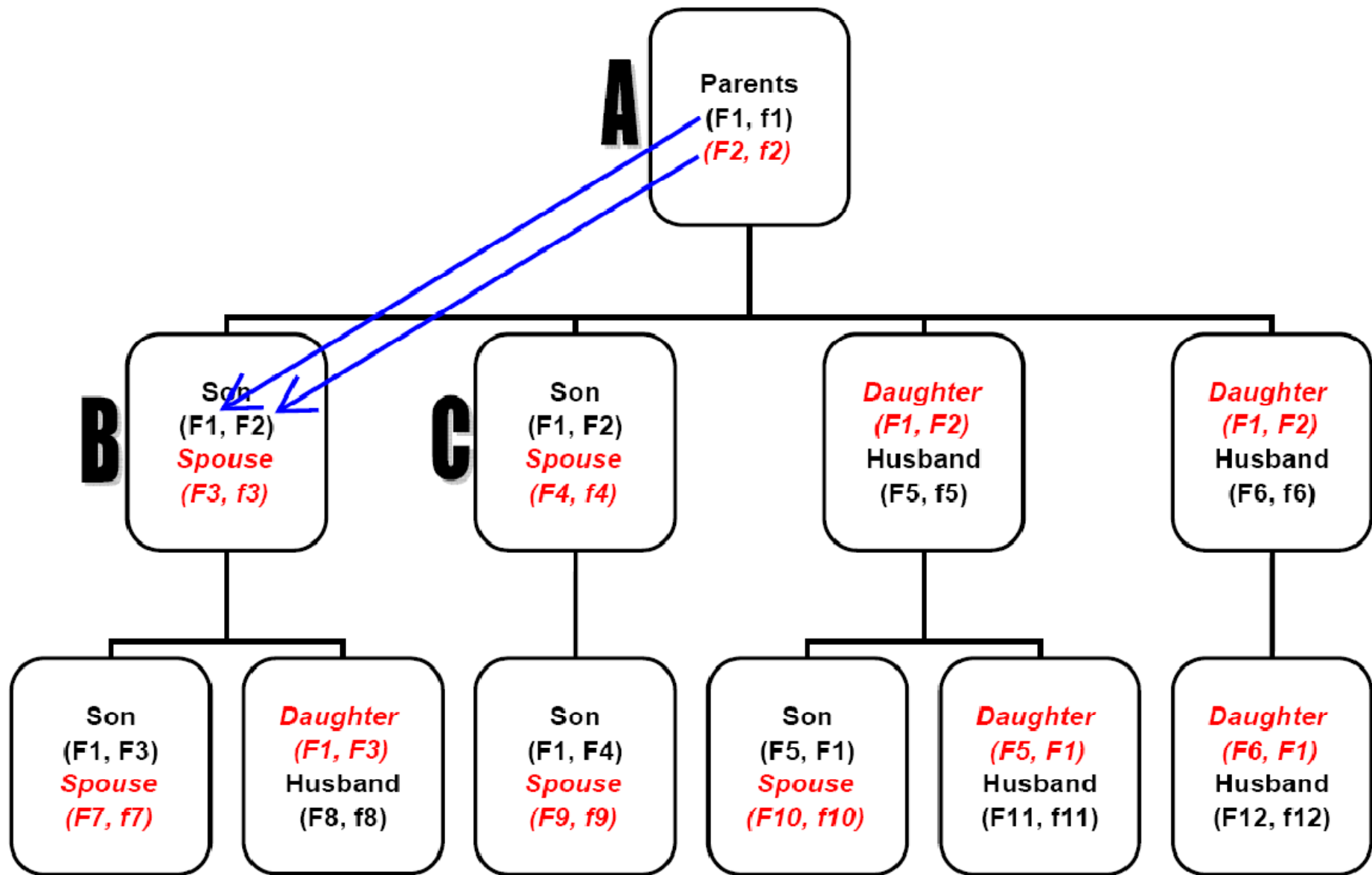


SURNAMES IN MEXICO

- Mexicans have two surnames: inherited from the father and mother's paternal lineage
- for example, former Mexican president Vicente Fox Quesada is identified by his given name (Vicente), his father's paternal name (Fox) and his mother's paternal name (Quesada)
- respondents were asked to provide the – (i) given name; (ii) paternal surname; (iii) maternal surname, for each household member
- hence couple headed households have four associated surnames
- [Figure 1: Family tree]



Figure 1: Family Tree



Notes: We use the convention that the head's surnames are written in standard (black) font, and those of his wife are written in (red) italics. Paternal surnames are indicated in upper case ($F1, F2$) and maternal surnames are indicated in lower case ($f1, f2$). First names are not shown as they are not relevant for the construction of extended family ties. Each household in the family tree is assumed to be couple headed purely to ease the exposition.

Table 1: The Number of Extended Family Links, by Type of Link

Couple Headed Households

Mean, standard error in parentheses clustered by village

Heads have more family ties present than their wives
Because women more mobile at the time of marriage
(Rosenzweig and Stark 1989)

Outside the Household and in the Village

	<u>Parent</u>		<u>Children Aged 0-16</u>		<u>Adult Children</u>		<u>Siblings</u>		<u>All</u>	
	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
From head of household to:	.461 (.010)	-	-	-	.652 (.066)	-	2.23 (.111)	-	3.34 (.164)	-
From spouse of household to:	.250 (.007)	-	-	-	.652 (.066)	-	1.63 (.103)	-	2.54 (.160)	-

Inside the Household

	<u>Parent</u>		<u>Children Aged 0-16</u>		<u>Adult Children</u>		<u>Siblings</u>		<u>All</u>	
	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
From head of household to:	.062 (.003)	.079 (.006)	3.23 (.027)	3.10 (.043)	.021 (.015)	.707 (.021)	.033 (.002)	.052 (.007)	4.15 (.035)	4.02 (.056)
From spouse of household to:	.018 (.001)	.021 (.003)	3.23 (.027)	3.10 (.043)	.821 (.015)	.787 (.021)	.013 (.001)	.017 (.003)	4.09 (.034)	3.93 (.054)

Notes: The sample is restricted to couple headed households in the baseline survey. Standard errors are clustered by village. Of the 22553 households that can be tracked in the first and third waves of *Progres*, 84.2% report to be couple headed in October 1997 (wave 1). We define the head of the household to be the male among the couple. By construction, the number of family links to parental households is always two conditional on such a family link existing. By construction, the number of children of the couple inside and outside the household are identical for the head and the spouse. Adult children are defined to be at least 17 years of age.

Similar family structures within the household (under the same roof)

EMPIRICAL METHOD

- *Progresa* research design allows for the evaluation of *Progresa* on eligible and non-eligible households
- under the twin assumptions of random assignment, and that control villages are not affected by the program:
 - eligibles in control villages provide a valid counterfactual for eligibles in treatment villages (TTE)
 - non-eligibles in control villages provide a valid counterfactual for non-eligibles in treatment villages (ITE)



EMPIRICAL METHOD

- core of the empirical analysis is to understand whether *TTE* and *ITE* are heterogeneous according to the presence and characteristics of extended family ties
- identification of heterogeneous *TTE* and *ITEs* relies on assumption that family structures are identical between (non) eligible households in treatment and control villages
- global structures similar across villages
- more heterogeneity across networks within a village in share of family that has primary or secondary school aged children, and amount of transfers the average household in the network is eligible for



Intra-generational links are more common than Inter-generational links

Table 3: Probability of an Extended Family Link

Couple Headed Households

Mean, standard errors in parentheses clustered by village

	Any Family Link (Connected)	Any Family Link of the Head	Any Family Link of the Spouse	Intra-generational Family Links				Inter-generational Family Links			
				Head to Head (Brothers)	Head to Spouse	Spouse to Head	Spouse to Spouse (Sisters)	Parents to Son	Parents to Daughter	Son to Parent	Daughter to Parent
Eligible Households											
Treatment	.817 (.011)	.693 (.012)	.550 (.014)	.506 (.013)	.351 (.015)	.338 (.014)	.306 (.013)	.150 (.009)	.077 (.006)	.169 (.008)	.108 (.007)
Control	.800 (.017)	.602 (.020)	.541 (.023)	.503 (.023)	.364 (.026)	.340 (.026)	.314 (.026)	.149 (.014)	.079 (.009)	.163 (.011)	.097 (.008)
Difference	.017 (.020)	.012 (.023)	.009 (.028)	.003 (.026)	-.013 (.030)	-.010 (.030)	-.008 (.030)	.002 (.016)	-.002 (.010)	.006 (.013)	.011 (.011)
Non-eligible Households											
Treatment	.808 (.016)	.712 (.018)	.523 (.021)	.532 (.021)	.332 (.018)	.258 (.016)	.248 (.017)	.198 (.015)	.112 (.010)	.142 (.014)	.074 (.008)
Control	.802 (.019)	.694 (.022)	.562 (.024)	.481 (.026)	.353 (.028)	.313 (.022)	.272 (.022)	.226 (.021)	.112 (.016)	.126 (.014)	.089 (.010)
Difference	.006 (.025)	.019 (.020)	-.039 (.032)	.051 (.034)	-.021 (.034)	-.055** (.027)	-.024 (.020)	-.028 (.026)	-.000 (.019)	.016 (.020)	-.015 (.013)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered by village. The sample is restricted to couple headed households that can be tracked over the first and third *Progres*a waves. Means and differences are reported for those households that have secondary school age children (aged 11 to 16) in the baseline survey of October 1997. The standard errors on the differences are calculated from running a corresponding OLS regression, which allows for the error terms to be clustered by village.

Similar family structures outside the household and within the village, across eligible and non-eligibles

FAMILY NETWORK DESCRIPTIVES

Family networks do not span more than three generations

Table 4: Family Network Descriptives

Means, standard deviation between villages in parentheses, standard deviation within villages in brackets

Treatment Villages	Size of Global Family Network	Network Size/Number of Households in Village	Diameter	Share that are Eligible	Share With Primary School Aged Children at Baseline	Share With Secondary School Aged Children at Baseline	Average Value of Potential Transfers Households in Network are Eligible for (March 1998, pesos)
Mean	7.66	.169	2.42	.518	.480	.480	746
Standard deviation between villages	(.249)	(.153)	(1.20)	(.233)	(.150)	(.139)	(181)
Standard deviation within villages	[.153]	[.153]	[2.14]	[.259]	[.262]	[.263]	[281]

Control Villages	Size of Global Family Network	Network Size/Number of Households in Village	Diameter	Share that are Eligible	Share With Primary School Aged Children at Baseline	Share With Secondary School Aged Children at Baseline	Average Value of Potential Transfers Households in Network are Eligible for (March 1998, pesos)
Mean	7.92	.163	2.51	.525	.484	.491	756
Standard deviation between villages	(.230)	(.141)	(1.14)	(.249)	(.145)	(.139)	(188)
Standard deviation within villages	[.155]	[.155]	[2.07]	[.229]	[.252]	[.258]	[297]

Notes: The sample is restricted to households that can be tracked over the first and third *Progreso* waves. There is one observation per family network so that each network has the same weight irrespective of the number of households within it. There are 1379 family networks in treatment villages covering 10569 households. There are 817 family networks in control villages covering 6471 households. The size of the network is the number of households in the network. The diameter of the networks is the longest distance between two households that exists in a network. We define two households that are directly connected to be of distance one to each other. Primary school aged children are defined to be those aged 6 to 10 and resident in the household. Secondary school aged children are defined to be those aged 11 to 16 and resident in the household. The average value of potential transfers households in the network are eligible for, are calculated among eligible households only. The standard deviations between and within villages take account of the fact that there are an unequal number of family networks in each village.

Lots of variation within the same village

OUTCOME: CHANGE IN SECONDARY SCHOOL ENROLMENT

- secondary school enrolment rate: fraction of 11-16 years olds in the household that are in full time schooling
- outcome variable: $\Delta Y_h = \text{change}$ in secondary school enrolment rate between wave 5 (November 1999) and wave 1 (October 1997)
- within household analysis – permanent unobserved family characteristics [preferences, ability, network structure] – are differenced out



Table 2: Descriptive Evidence on Enrolment Rates

Couple Headed Households

Mean, standard errors in parentheses clustered by village

Secondary School Enrolment Rates (children aged 11 to 16)

Eligibles, by Village Type

	Control	Treatment	All Households
	October 1997	October 1997	Difference in Difference
All children	.651 (.016)	.654 (.012)	.069*** (.016)
Boys	.681 (.017)	.685 (.013)	.043** (.022)
Girls	.603 (.018)	.612 (.014)	.102*** (.022)

Previously established results
last lecture (Shultz 2004)

Table 2: Descriptive Evidence on Enrolment Rates

Couple Headed Households

Mean, standard errors in parentheses clustered by village

Secondary School Enrolment Rates (children aged 11 to 16)

	<u>Eligibles, by Family Link Type</u>			
	Connected	Isolated	Connected	Isolated
	October 1997	October 1997	Difference in Difference	Difference in Difference
All children	.653 (.011)	.654 (.015)	.083*** (.017)	.001 (.031)
Boys	.681 (.012)	.698 (.016)	.044* (.023)	.040 (.047)
Girls	.608 (.012)	.610 (.019)	.131*** (.024)	-.030 (.044)
Similar enrolment rates at baseline				

No response among **eligible but isolated** households, neither for boys nor girls

Table 2: Descriptive Evidence on Enrolment Rates

Couple Headed Households

Mean, standard errors in parentheses clustered by village

Primary School Enrolment Rates (children aged 6 to 10)

Eligibles, by Family Link Type

	Connected	Isolated	Connected	Isolated
	October 1997	October 1997	Difference in Difference	Difference in Difference
All children	.927 (.005)	.890 (.013)	.013 (.012)	-.011 (.021)
Boys	.932 (.006)	.887 (.014)	.020 (.015)	-.001 (.031)
Girls	.925 (.007)	.901 (.012)	.005 (.015)	-.015 (.026)

Similar enrolment rates at baseline

Such high enrolment rates at baseline imply that conditional cash transfers for primary enrolment act as **de facto unconditional pure income transfers**

POLICY IMPLICATION

- interplay between the design of conditional cash transfer programs, the presence of extended family members, and household responses to the program
- if families share resources, how connected households respond to such policies on any given margin, will generally depend on the eligibility status of others in their network
- ignoring the presence and characteristics of the extended family can lead to an incomplete understanding of the forces driving the behavioral responses of households to large scale policy interventions in developing country settings
- resources spent on *Progresa* could be more efficiently targeted if the policy aim is to increase secondary school enrolment: revenue neutral policy alternatives [Todd and Wolpin 2006, Attanasio *et al* 2005]:
 - then expect *both* isolated and connected households to increase secondary enrolment



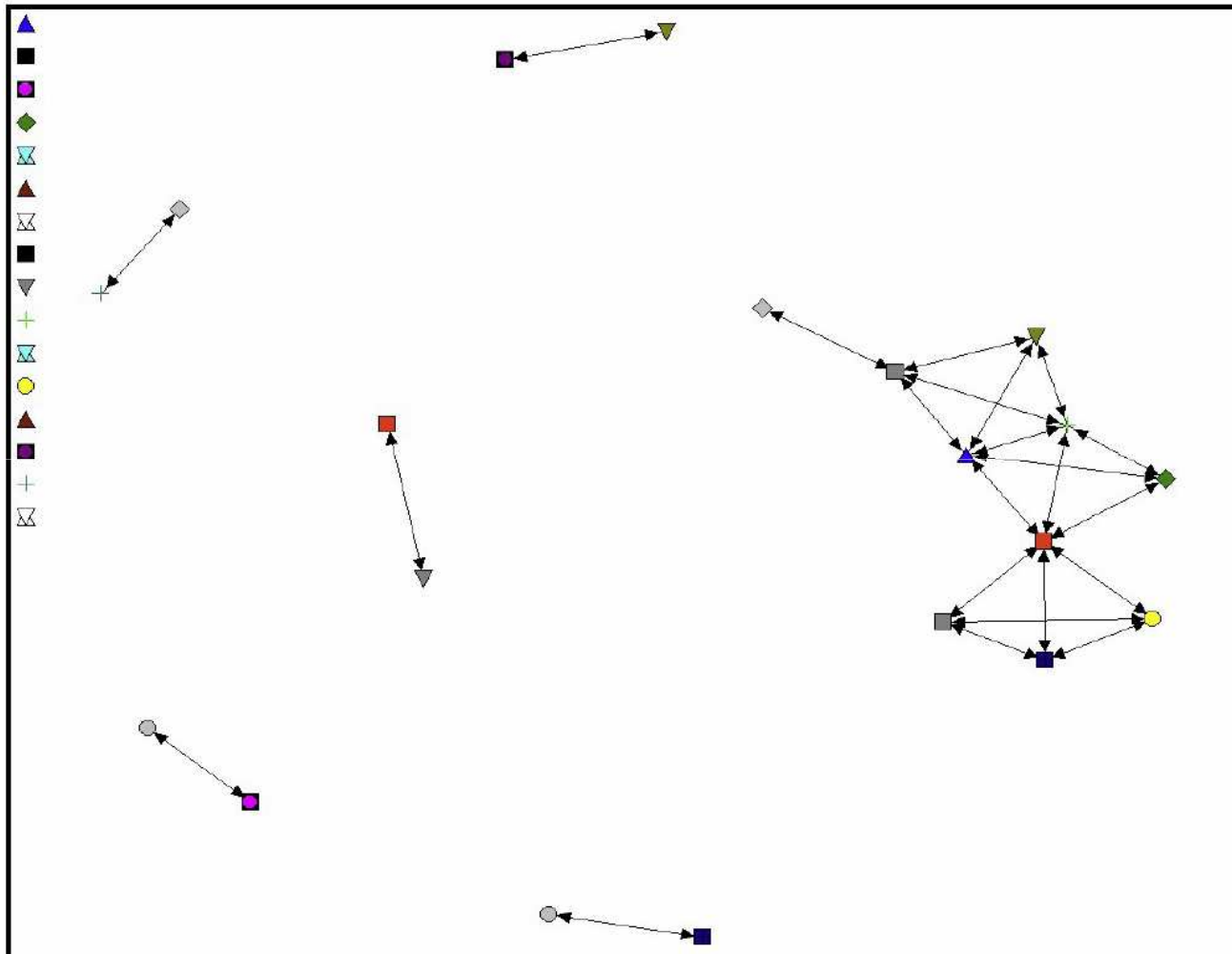
EXTENDED FAMILY NETWORKS ACROSS VILLAGES

- Previous study focuses on heterogeneous policy responses depending on presence and characteristics of extended family network
- Took network structures as **given**
- Useful to explore correlations at **village level** between network structures and village level characteristics



Figure 4: Family Network Graphs, at Median Village Size

A. Disperse Village

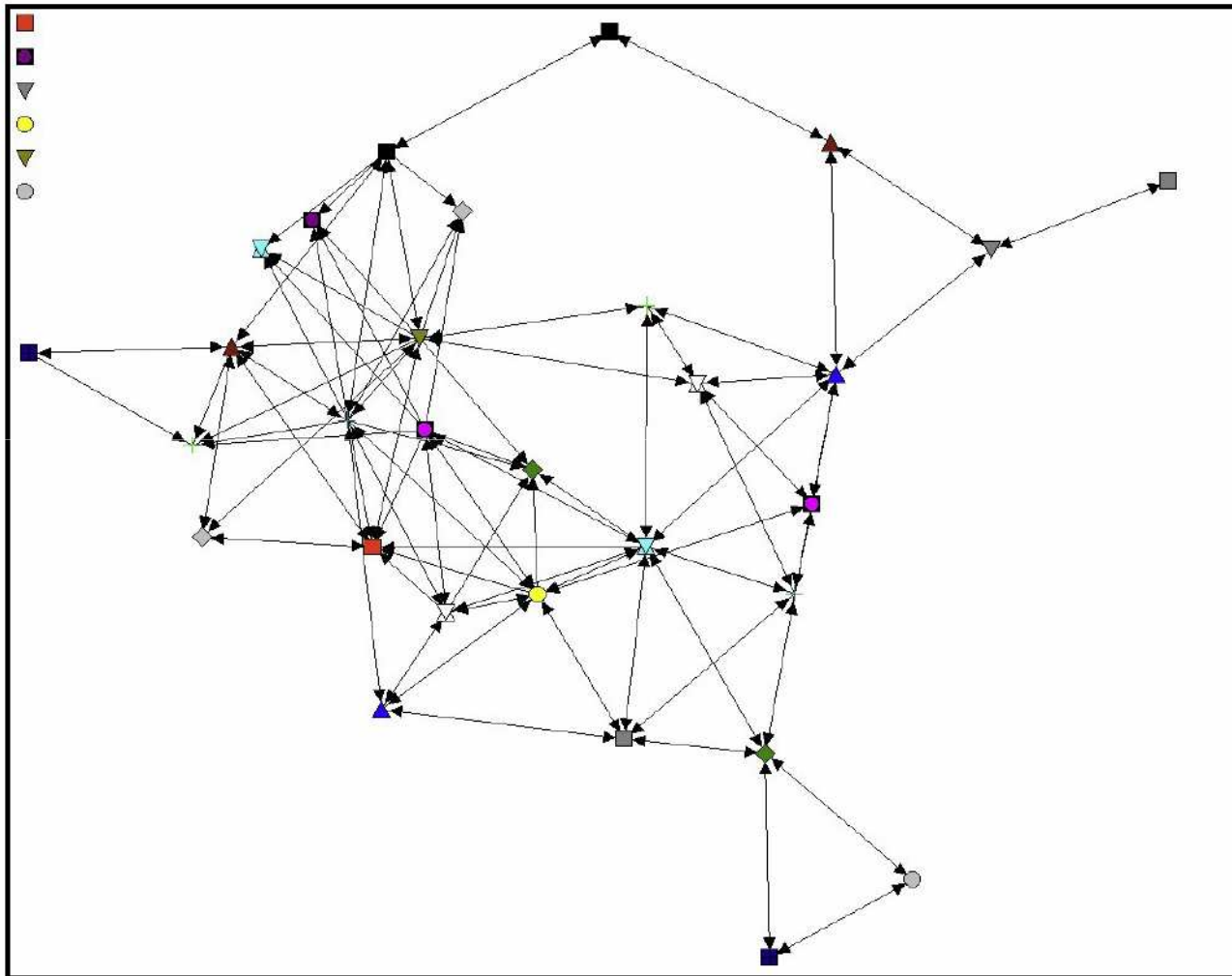


Notes: The two villages shown in Figures A and B have the same number of households in them. The number of households in each is 36, which is the median village size in the Progres data. Each node represents a household. Each link between households correspond either to a parent/child link, a child/parent link, or a sibling link. Single node households that are not linked to any other households are shown in the top left hand corner of each graph. The figures are generated using UCINET.



Figure 4: Family Network Graphs, at Median Village Size

B. Interconnected Village



Notes: The two villages shown in Figures A and B have the same number of households in them. The number of households in each is 36, which is the median village size in the Progresa data. Each node represents a household. Each link between households correspond either to a parent/child link, a child/parent link, or a sibling link. Single node households that are not linked to any other households are shown in the top left hand corner of each graph. The figures are generated using UCINET.



Empirical Evidence: Extended Family Networks

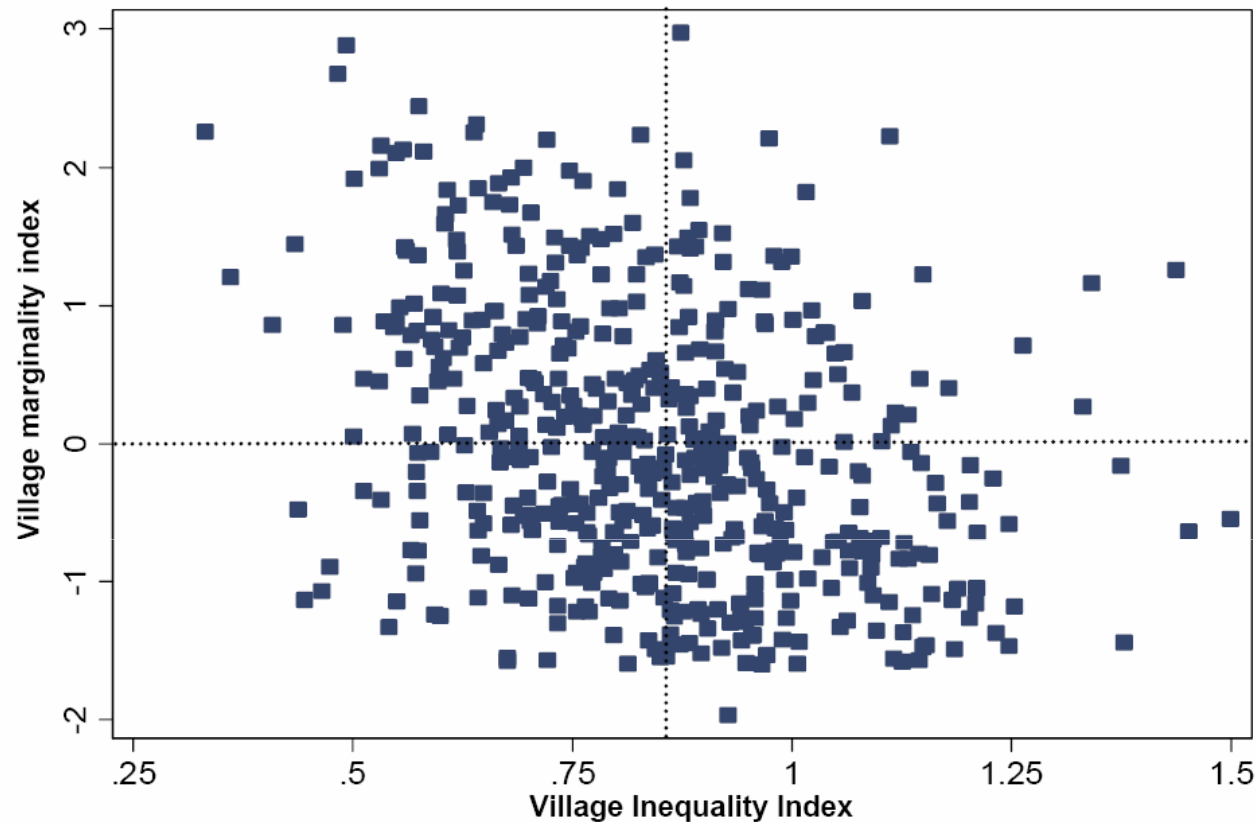
- estimate an OLS regression at the village level of the following form,

$$N_v = \beta M_v + \gamma Q_v + \delta S_v + u_v,$$

- where N_v = number of family networks in village v , M_v = village marginality index; Q_v = village inequality index, and S_v = village size
 - absent any exogenous variation in the marginality of the village or inequality within it, it is impossible to identify whether network structures endogenously respond to these characteristics of the village economy, or whether the features of the local economy are themselves shaped by the nature of extended family networks within them
 - [Table 5, Cols 1-3]
 - [Table 5, Cols 4-5]
-

Figure 1: The Village Economy

A: Village Marginality and Inequality



Notes: In each figure, there is one observation at the village level. The village marginality index is constructed from information on the share of illiterate adults in the village, the share of dwellings without water, drainage systems, electricity, and with floors of dirt, the average number of occupants per room in village households, the share of the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare index is a weighted average of household income (excluding children), household size, durables, land and livestock, education, and other physical characteristics of the dwelling. The index is designed to give relatively greater weight to correlates of permanent income rather than current income. An increase in the index implies the households is less poor. The measure of village inequality is the standard deviation of the welfare index of all households in the village. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The household welfare index is defined relative to a state norm. Hence the village inequality index is standardized at the state level. Hence the village marginality index and standard deviation of household welfare at the village level are both standardized at the regional level. Two villages are dropped (of size two and eleven) in which there are no households with any extended family links to others in the same village.



Table 5: Family Network Structures and Village Characteristics

OLS regression estimates, robust standard errors in parentheses

Dependent Variable:	Village Level Statistics				
	Number of Family Networks			Size of the Largest Family Network	Share of Village Households that are in the Largest Family Network
	(1)	(2)	(3)	(4)	(5)
Village marginality index	-.237* (.129)	-.253** (.124)	-.406** (.164)	1.67** (.809)	.029** (.014)
Village inequality index	2.27*** (.559)	2.01*** (.582)	1.79** (.718)	-11.1*** (3.73)	-.219*** (.071)
Village size	.339*** (.721)	.346*** (.736)	.347*** (.770)	.718*** (5.40)	.230*** (.045)
State fixed effects	No	Yes	No	No	No
Municipality fixed effects	No	No	Yes	Yes	Yes
R-squared	.167	.197	.243	.749	.325
Observations	504	504	504	504	504

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. There is one observation for each village, and each dependent variable is constructed using couple headed households that can be tracked over the first and third Progesa waves in each village. The village marginality index is constructed from information on the share of illiterate adults in the village, the share of dwellings without water, drainage systems, electricity, and with floors of dirt, the average number of occupants per room in village households, the share of the population working in the primary sector, distances from other villages, and health and school infrastructures located in the village. A higher marginality index corresponds to the village being more marginal (poorer). The household welfare index is a weighted average of household income (excluding children), household size, durables, land and livestock, education, and other physical characteristics of the dwelling. The index is designed to give relatively greater weight to correlates of permanent income rather than current income. An increase in the index implies the households is less poor. The measure of village inequality is the standard deviation of the welfare index of all households in the village. Village size is defined as the number of households in the village. The village marginality index is standardized across all villages. The household welfare index is defined relative to a state norm. Hence the village inequality index is standardized at the state level. Two villages are dropped (of size two and eleven) in which there are no households with any extended family links to others in the same village. The villages in the sample cover 7 states and 115 municipalities. Robust standard errors are reported.



SURVEY DESIGN

- have documented that 20% of couple headed households are isolated in that none of their extended family members are geographically proximate in the same village
- incidence of this type of isolatedness is therefore at least as high as the incidence of single headedness, which affects 15% of households in our data
- the importance of designing future surveys to identify isolated households in other settings, and more generally, to establish the social ties between households in survey data [Conley and Udry 2005, Akresh 2005, Dercon et al 2005]



EMPIRICAL EXAMPLES

- Progresa and extended family networks in Mexico
- The Women's Group Programme in Malawi and networks of brothers and sisters
- The Ultra-Poor Programme in Bangladesh and village networks



MOTIVATION

- Households in rural areas of developing countries face a wide variety of risks and adverse events:
 - agriculture, employment, health...
- Limited scope to cope with adverse events through formal channels:
 - Few government programmes exist; concerns over how they are targeted
 - Insurance and credit not widely available
- Households might engage in *informal* strategies to cope with adverse events
 - Transfers, loans, gifts, labour sharing, etc (Besley, 1995)
- Social ties important for such informal strategies



MOTIVATION

- Interventions and policies may interact with the informal risk sharing provided by one's social ties
 - They may crowd out such risk sharing
 - Interventions targeted at specific portions of a community, such as women or the poorest households, can make these groups more attractive to transact with and so improve risk sharing
- Theoretical literature in economics suggests that there may also be an *optimal network size* within which risk can be shared (Genicot and Ray 2003):
 - Beyond a certain size, smaller groups within the larger social network may decide to walk away from the arrangement and share risk only with each other
 - In large groups, easy to free ride – expect other social ties to provide help



RESEARCH QUESTION

- Objective: Understand how a **Women's Group Intervention** in rural Malawian communities changed risk sharing arrangements in extended family networks
- Intervention changed social interactions in the treated communities, which may improve risk sharing
- Likely to have interacted with existing risk sharing networks



INTERVENTION

- Women's group intervention implemented by Mai Mwana in Mchinji District, in the Central region of Malawi
- Set up by the Institute of Child Health at UCL. Similar interventions implemented in Malawi, Nepal, India, and Bangladesh
- A facilitator organises fortnightly meetings in the village to improve reproductive health (i.e. during pregnancy, delivery, and post partum)
- The groups are encouraged to follow a participatory approach: identify problems, devise strategies to overcome them, and try to involve the wider community in their implementation



EXPERIMENTAL DESIGN

- Participation in groups is voluntary (30% average participation rate)
- Intervention started in 2005
- Intervention set up as part of a cluster randomised control trial:
 - 12 intervention clusters; 12 control
 - Each cluster contains ~14 villages
 - Each village contains ~ 42 households with a woman of reproductive age



DATA

- Survey of randomly selected women of reproductive age (independently of group participation)
- Two waves collected: Oct 2008-Feb 2009; Oct 2009-Feb 2010
 - Both waves are collected after the intervention started
- PDA based: increased accuracy, acceptance and interviewer motivation



DATA

- Household consumption, including non-purchased
 - Equivalence of non-standard units obtained through visits to markets
- Adverse events:
 - Crop loss, business, theft
- 2 measures of intensity:
 - Dummy variable: Whether or not household experienced crop loss
 - Relative importance of loss to household: estimated loss as a fraction of estimated (pre-intervention) monthly consumption



FAMILY ARRANGEMENTS

- Main ethnic group in the area, Chewa are matrilineal
 - Man moves to woman's village after marriage
 - Both can move to the man's village if the man pays the woman's family a marriage payment (relatively common in Mchinji because the Chewa have integrated with a patrilineal ethnic group – the Ngoni)
- A woman's maternal uncle holds the power in the family
 - He has general responsibility for the welfare of the family, and settles internal disputes, and obtains land for the family to use among other things
 - A woman's eldest brother would have this responsibility for her children



DESCRIPTION OF FAMILY NETWORKS

	Any Family Link	Any Family Link of Head	Any Family Link of Spouse	Any Intragenerational Links		Any Inter-generational Links			
				Head Mother	Spouse Mother	Head		Spouse	
						Brothers	Sisters	Brothers	Sisters
Alive									
Treated Zones	0.997	0.988	0.964	0.815	0.822	0.858	0.847	0.89	0.869
	[0.002]	[0.003]	[0.011]	[0.033]	[0.015]	[0.026]	[0.03]	[0.018]	[0.017]
Control Zones	0.995	0.984	0.979	0.785	0.875	0.857	0.837	0.885	0.887
	[0.003]	[0.007]	[0.007]	[0.022]	[0.015]	[0.022]	[0.021]	[0.015]	[0.017]
Difference	0.002	0.004	-0.015	0.03	-0.054**	0	0.01	0.005	-0.018
	[0.003]	[0.007]	[0.013]	[0.039]	[0.021]	[0.033]	[0.037]	[0.023]	[0.024]
Living in the Same Village									
Treated Zones	0.825	0.723	0.704	0.434	0.511	0.426	0.38	0.328	0.368
	[0.036]	[0.041]	[0.032]	[0.032]	[0.025]	[0.064]	[0.04]	[0.024]	[0.042]
Control Zones	0.809	0.684	0.708	0.44	0.563	0.408	0.378	0.288	0.346
	[0.034]	[0.026]	[0.039]	[0.03]	[0.032]	[0.048]	[0.027]	[0.024]	[0.045]
Difference	0.016	0.039	-0.004	-0.006	-0.052	0.018	0.003	0.04	0.023
	[0.049]	[0.047]	[0.05]	[0.043]	[0.04]	[0.079]	[0.048]	[0.034]	[0.061]

ESTIMATION EQUATION

- We focus on informal risk sharing after crop losses
 - Estimate whether households affected by a crop loss shock protect their consumption better if they live in a women's group village than in a control village

$$\Delta \log c_{hvt} = \alpha \Delta \text{crop}_{hvt} + \gamma \Delta \text{crop}_{hvt} * D + \Delta X_{hvt} + v_t + \Delta \varepsilon_{hvt}$$

- D = 1 if in women's group village, 0 if in control
- Consider also how the effects of the intervention on informal risk sharing vary depending on the size of one's family network

$$\Delta \log c_{hvt} = \alpha \Delta \text{crop}_{hvt} + \mu \Delta \text{crop}_{hvt} * D + \sum_{i=1}^3 \{ \alpha_i \mathbf{1}(N_{ihvt} = 1) + \gamma_i \mathbf{1}(N_{ihvt} = 1) * \Delta \text{crop}_{hvt} + \beta_i \mathbf{1}(N_{ihvt} = 1) * \Delta \text{crop}_{hvt} * D \} + \Delta X_{hvt} + v_t + \Delta \varepsilon_{hvt}$$

- $\mathbf{1}(\cdot)$ is an indicator function, which =1 if $N_{ihvt} = 1$; 0 otherwise
 - $N_{1hvt} = 1$ if hhld has 0 relatives of a particular type
 - $N_{2hvt} = 1$ if hhld has 1 or 2 relatives of a particular type
 - $N_{3hvt} = 1$ if hhld has 3 or more relatives of a particular type



RESULTS – BASIC SPECIFICATION

	$\Delta \text{Ln(Cons)}$		$\Delta \text{Ln(Food)}$	
	Crop=1 or 0	Crop=Loss/ Pred. Cons	Crop=1 or 0	Crop=Loss/ Pred. Cons
Δcrop	-0.09* (0.05) [0.09]	-0.0592** (0.022) [0.042]	-0.098 (0.06) [0.12]	-0.0485* (0.028) [0.042]
$\Delta \text{crop} * \text{D}$	0.13* (0.06) [0.08]	0.0411* (0.023) [0.06]	0.138* (0.078) [0.07]	0.0308 (0.029) [0.06]
N	1245	1221	1243	1219



RESULTS – BY NETWORK SIZE

	[1] Δtot_cons	[2] Δtot_cons	[3] Δtot_cons	[4] Δtot_cons
	Sisters Alive		Brothers Alive	
	Crop = 1 or 0	crop = Loss/Pred. Cons	Crop = 1 or 0	crop = Loss/Pred. Cons
Δcrop	-0.119 [0.0825]	-0.0713 [0.0473]	0.0832 [0.0644]	-0.00641 [0.0254]
Δcrop*D	0.13 [0.118]	0.0569 [0.0487]	-0.0365 [0.0879]	-0.0175 [0.0261]
N _{1hvt} = 1	0.0265 [0.0637]	0.0311 [0.0768]	-0.089 [0.0969]	-0.0376 [0.0827]
N _{1hvt} = 1*Δcrop	-0.0443 [0.0960]	-0.0644 [0.185]	-0.0838 [0.154]	0.0847 [0.0624]
N _{1hvt} = 1*D	0.0262 [0.0856]	0.00617 [0.0966]	0.00387 [0.164]	-0.0576 [0.150]
N _{1hvt} = 1*Δcrop*D	0.0542 [0.204]	0.176 [0.232]	0.219 [0.221]	-0.0307 [0.0921]
N _{3hvt} = 1	-0.0518 [0.0489]	-0.0649 [0.0411]	-0.127** [0.0577]	-0.117* [0.0662]
N _{3hvt} = 1*Δcrop	0.0704 [0.0990]	0.0221 [0.0539]	-0.403*** [0.102]	-0.242*** [0.0478]
N _{3hvt} = 1*D	0.103 [0.0600]	0.124** [0.0536]	0.0598 [0.0808]	0.0641 [0.0861]
N _{3hvt} = 1*Δcrop*D	-0.00346 [0.152]	-0.0552 [0.0647]	0.388*** [0.114]	0.284*** [0.0564]
Observations	1,238	1,214	1,235	1,211
R-squared	0.415	0.419	0.429	0.433



RESULTS

- Importance of mother's brothers
 - In line with family arrangements in this zone of Malawi
- In control areas, risk sharing is worse for those mothers with 3 or more brothers than those that have 1 or 2
 - In line with Genicot and Rey (2003) prediction that a larger network might be detrimental (free-rider problem)
- In treatment areas, this is not the case. The program must help by either:
 - Reducing the free-rider problem by making more costly to free-ride (knowledge someone is not being helped, shame...)
 - Or facilitating new risk sharing arrangements substituting which substitute for the malfunctioning network



EMPIRICAL EXAMPLES

- Progresa and extended family networks in Mexico
- The Women's Group Programme in Malawi and networks of brothers and sisters
- The Ultra-Poor Programme in Bangladesh and village networks

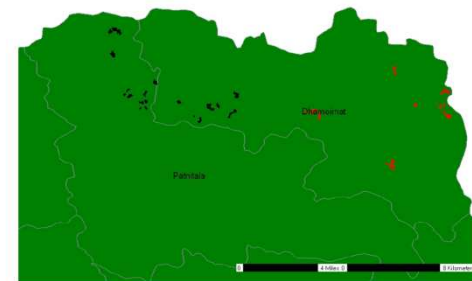
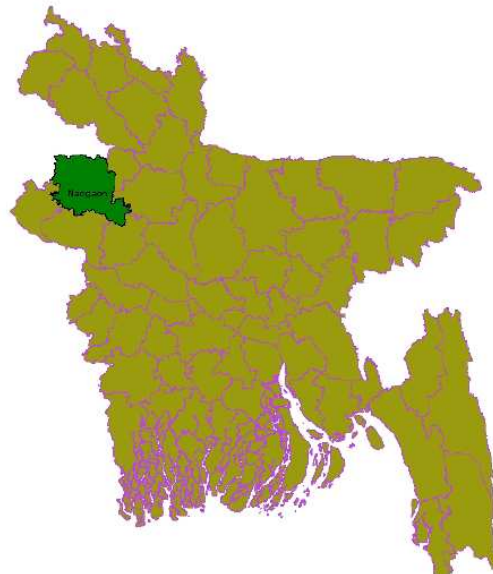


BANGLADESH ULTRA-POOR PROGRAM

Ultra-poor Networks

- ▶ Using subsample data to analyze the effects of the ultrapoor program on network characteristics
- ▶ In one subdistrict (Naogaon), we chose a number of spots to survey all households and map their entire network
- ▶ We have 3 survey waves (2008, 2009 2010), 2011 data is on the way

Naogaon – Location of surveyed households

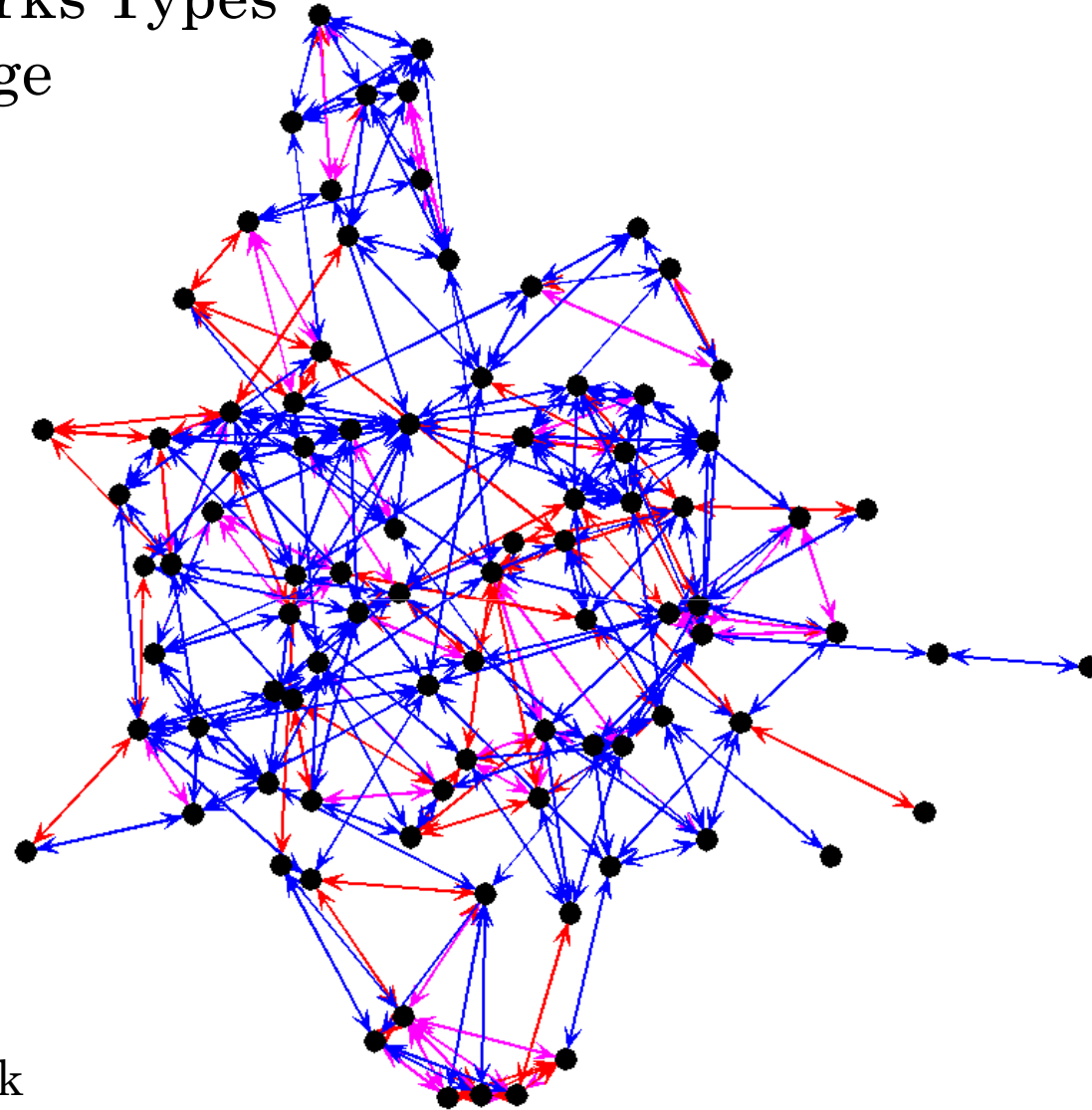


Networks at Baseline

- ▶ We define 4 types of aggregate networks:
 1. Family network – All first-degree family members (parents, siblings, parents and sibling in law, children)
 2. Market network – Labor, credit and asset market transactions
 3. Insurance network – Food exchange partners and last one year's transfer transactions
 4. Economic network – combination of market and insurance networks



Multiple Networks Types From One Village



Red = family link
Blue = economic link
Purple = both types of link



Table 1: Characteristics of Family Networks, by Treatment Group and Wave
Means, standard deviations in parentheses

	Wave 1 (Baseline)			Wave 3			Percentage Change between Waves		
	(1) Treated	(2) Control	(3) t-test	(5) Treated	(5) Control	(6) t-test	(7) Treated	(8) Control	(9) t-test
Number of Households in Spot	89.20 (12.22)	79.18 (5.49)	3.13***	86.90 (13.53)	78.73 (4.71)	2.44**	-2.58%	-0.57%	.3992
Number of Distinct Components	25.70 (9.26)	25.45 (6.91)	.0836	24.50 (6.27)	23.82 (7.70)	0.2514	-4.67%	-6.43%	-.1091
Number of Links	100.10 (27.00)	89.64 (23.40)	1.13	94.05 (23.84)	90.18 (25.12)	0.4176	-6.04%	0.61%	.5029
Network Link Density	.0259 (.0080)	.0289 (.0062)	-1.16	.0257 (.0061)	.0295 (.0077)	-1.41	-.77%	2.08%	.2142
Network Clustering Coefficient	0.62 (.1064)	0.65 (.1043)	-.6472	.5691 (.1039)	.6039 (.1203)	-.8079	-8.69%	-6.92%	.1589
Size of Maximal Clique	5.00 (.7862)	5.27 (.7897)	-.9215	4.75 (.8507)	5.18 (1.1677)	-1.07	-5.00%	-1.76%	.3160
Diameter of Largest Component	7.25 (3.48)	7.91 (3.00)	-.5524	8.15 (2.34)	5.64 (2.87)	2.48**	12.41%	-28.69%	-2.03**
Observations (Spots)	20	11		20	11		20	11	

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. In Columns 1 to 3 the Wave 1 survey period runs from Q4 2007 to Q1 2008. In Columns 4 to 6 the Wave 3 period covers Q1 2010. Columns (3) and (6) provide the t-statistic for difference in means. Column (9) tests the difference-in-difference. The number of components is the maximum number of groups households in the spot can be split into such that no households in different groups are connected. The number of links counts only links to households that were in the spot in the baseline. Network link density is the ratio of actual links to potential links in the network. The network clustering coefficient is the ratio of triangles to connected triples in the graph. A clique is a group of households such that each households in the group is connected to all others in the group. The diameter of a component is the longest path between a pair of nodes in the component.



Table 2: Characteristics of Economic Networks, by Treatment Group and Wave
Means, standard deviations in parentheses

	Wave 1 (Baseline)			Wave 3			Percentage Change between Waves		
	(1) Treated	(2) Control	(3) t-test	(5) Treated	(5) Control	(6) t-test	(7) Treated	(8) Control	(9) t-test
Number of Households in Spot	89.20 (12.22)	79.18 (5.49)	3.13***	86.90 (13.53)	78.73 (4.71)	2.44**	-2.58%	-0.57%	0.3992
Number of Distinct Components	2.95 (6.27)	2.64 (7.70)	.1143	3.70 (1.45)	3.36 (.92)	.7968	25.42%	27.27%	-.0109
Number of Links	240.35 (23.84)	211.91 (25.13)	3.07***	227.90 (45.63)	229.09 (29.84)	-.0875	-5.18%	8.11%	1.80*
Network Link Density	.0560 (.0061)	.0636 (.0077)	-2.82***	.0548 (.0113)	.0676 (.0051)	-4.32***	-2.14%	6.29%	1.30
Network Clustering Coefficient	.2955 (.1039)	.2711 (.1203)	.5665	.2672 (.0351)	.2532 (.0353)	1.06	-9.58%	-6.60%	.2308
Size of Maximal Clique	4.65 (.8507)	4.55 (1.1677)	.2499	4.55 (.5104)	4.54 (.5222)	.0514	-2.15%	-0.22%	.2023
Diameter of Largest Component	7.15 (2.34)	6.55 (2.87)	.5933	7.30 (1.03)	6.36 (.67)	3.07***	2.10%	-2.90%	-.3218
Observations (Spots)	20	11		20	11		20	11	

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. In Columns 1 to 3 the Wave 1 survey period runs from Q4 2007 to Q1 2008. In Columns 4 to 6 the Wave 3 period covers Q1 2010. Columns (3) and (6) provide the t-statistic for difference in means. Column (9) is tests the difference-in-difference. The number of components is the maximum number of groups households in the spot can be split into such that no households in different groups are connected. The number of links counts only links to households that were in the spot in the baseline. Network link density is the ratio of actual links to potential links in the network. The network clustering coefficient is the ratio of triangles to connected triples in the graph. A clique is a group of households such that each households in the group is connected to all others in the group. The diameter of a component is the longest path between a pair of nodes in the component.



CONCLUSIONS

- Empirical examples illustrate how:
 - How networks aid risk sharing or resource sharing
 - How multiple network types overlap
 - Networks might themselves respond to interventions

