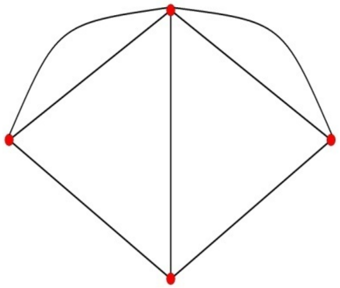
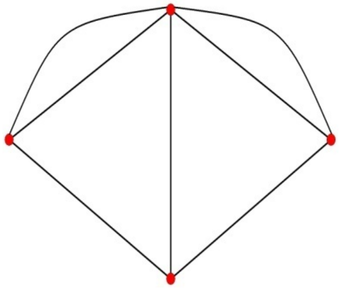
Social and Economic Networks



Matthew O. Jackson

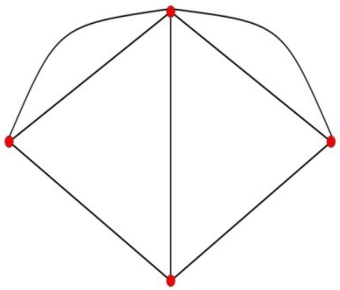


Lecture 1
Social and Economic Networks: Background

Matthew O. Jackson
NBER July 22, 2014

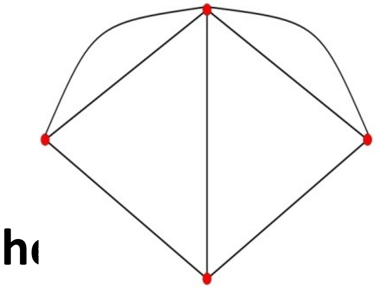
www.stanford.edu\~jacksonm\Jackson-NBER-slides2014.pdf

Lecture 1



- Crash course in some basic social network background

Why Study Networks?

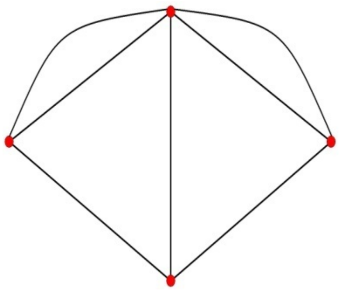


- **Many economic, political, and social interactions are shaped by the structure of relationships**:
  - trade of goods and services, most markets are not centralized!…
  - sharing of information, favors, risk, ...
  - transmission of viruses, opinions...
  - access to info about jobs...
  - choices of behavior, education, new technologies,...
  - political alliances, trade alliances…

- **Social networks influence behavior**
  - crime, employment, human capital, voting, product adoption, …
  - networks *exhibit heterogeneity, but have underlying structures that we can measure and model and use to understand implications for behavior, welfare, and policy*

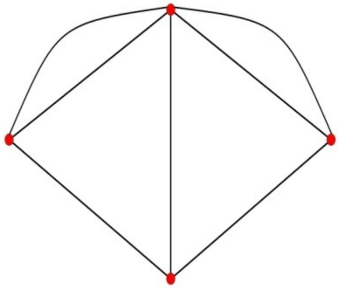
- **Pure interest in social structure**
  - understand social network structure

Synthesize



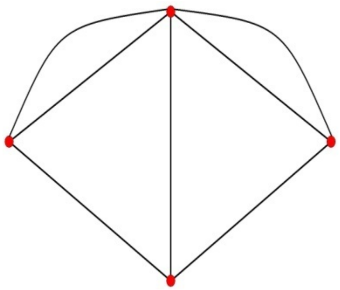
- Many literatures deal with networks
  - Sociology
  - Economics
  - Computer Science
  - Statistical Physics
  - Math (random graph)...
- What have we learned?
- What is the state of the art?
- What are important areas for future research?

Four parts:



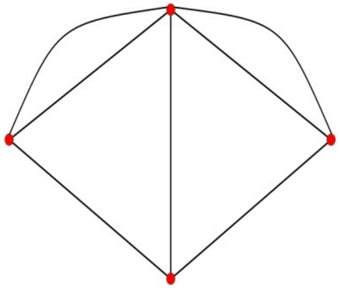
- Background, basics, some measures,
- Peer effects, identification
- Diffusion, more on identification, estimation
- Transmission of shocks…

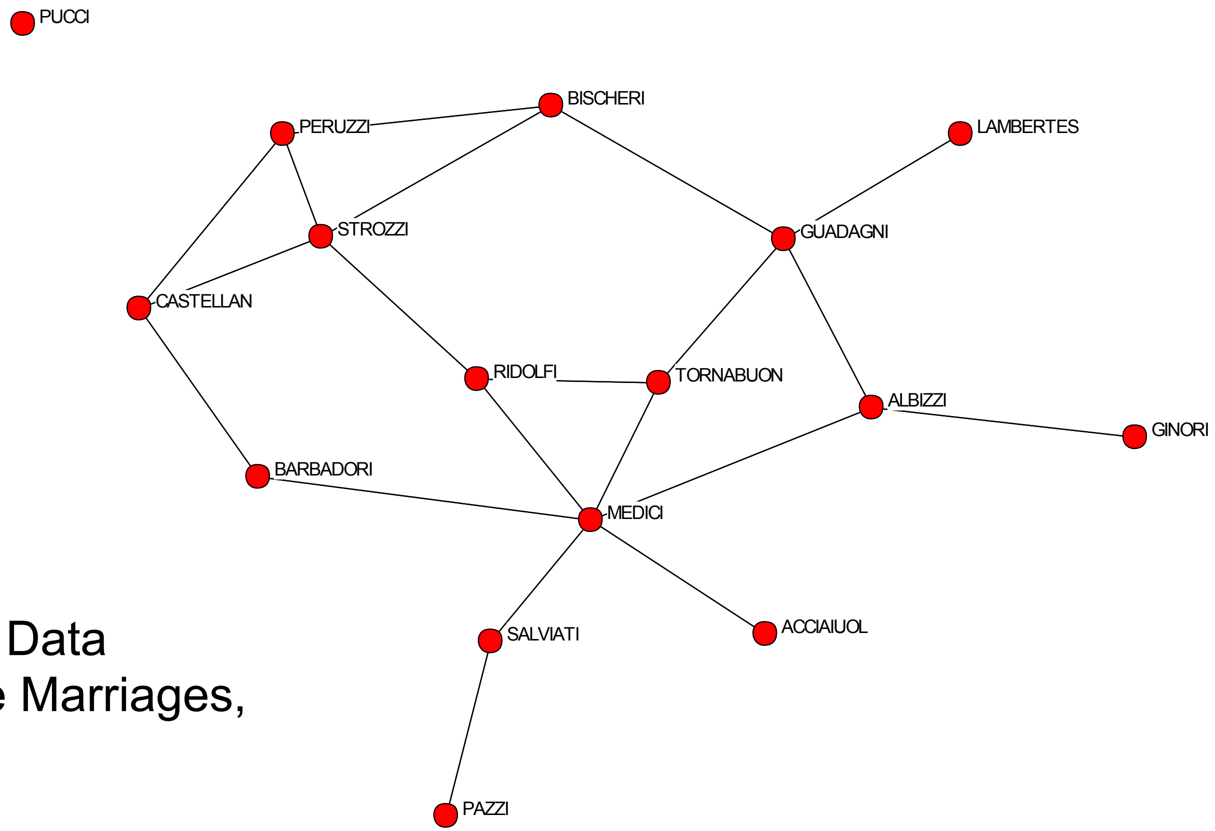
A Few Examples of Networks



- Idea of some data
- View of a few applications
- Preview some questions

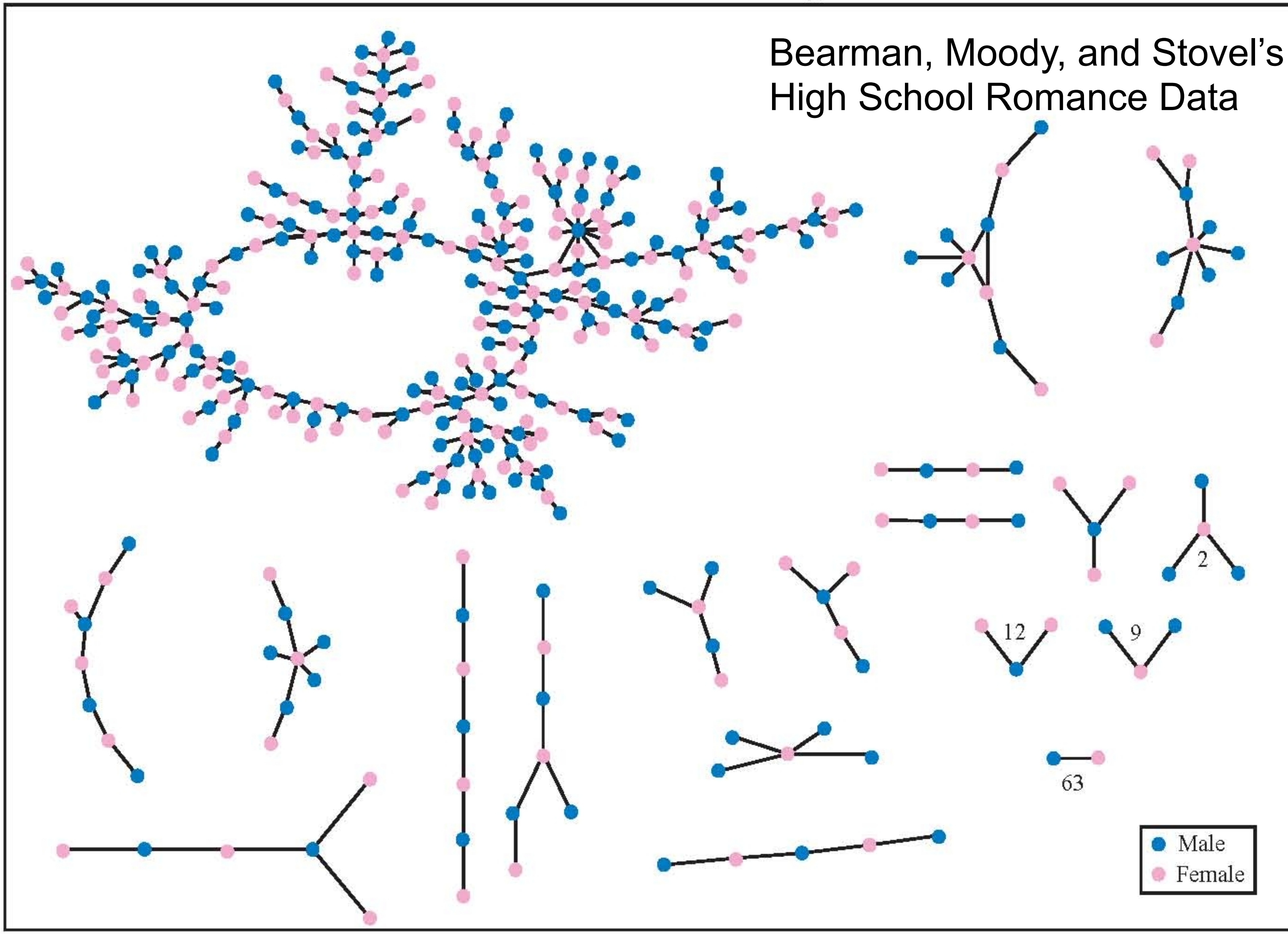
Examples of Social and Economic Networks



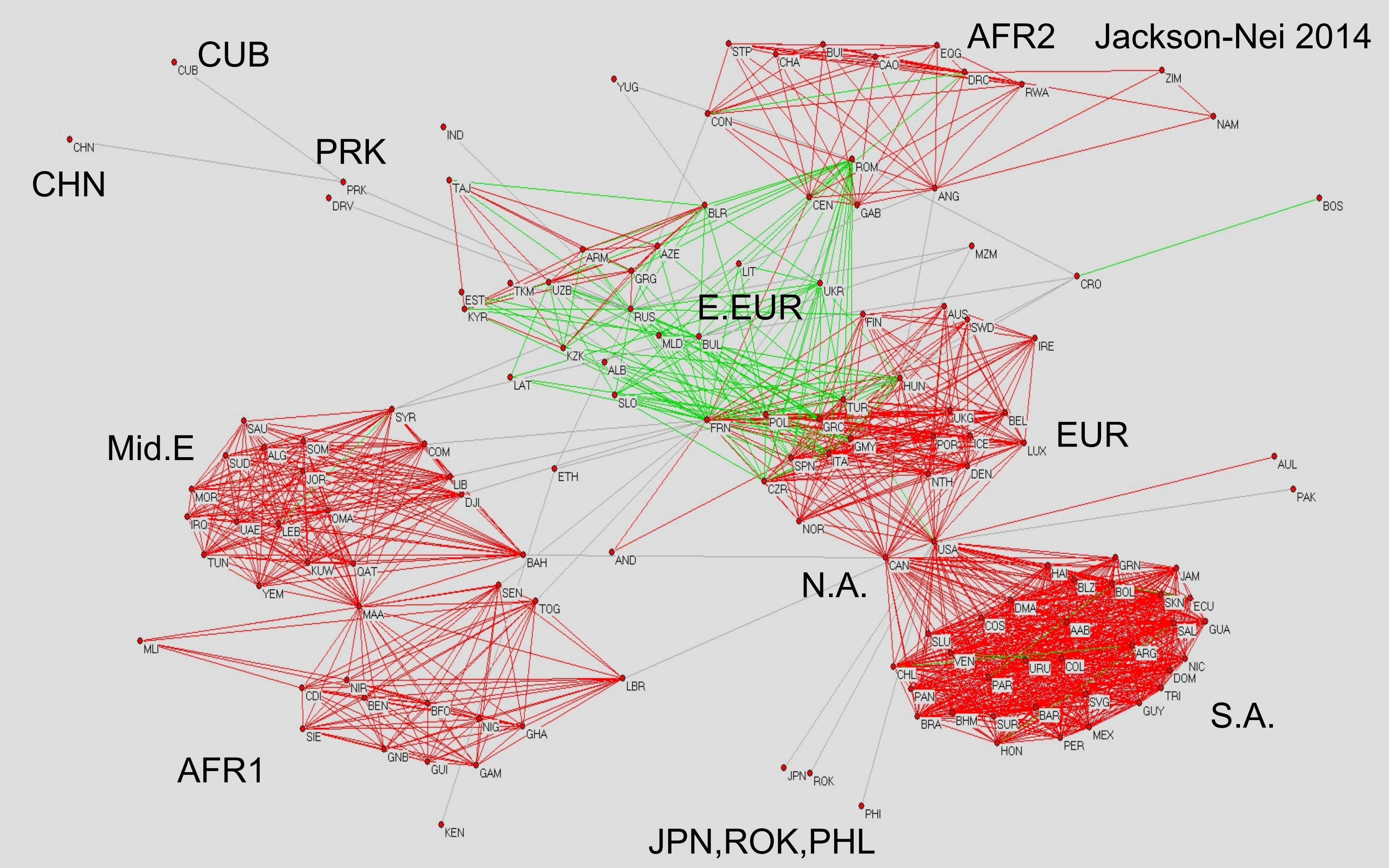


Padgett's Data
Florentine Marriages,
1430's

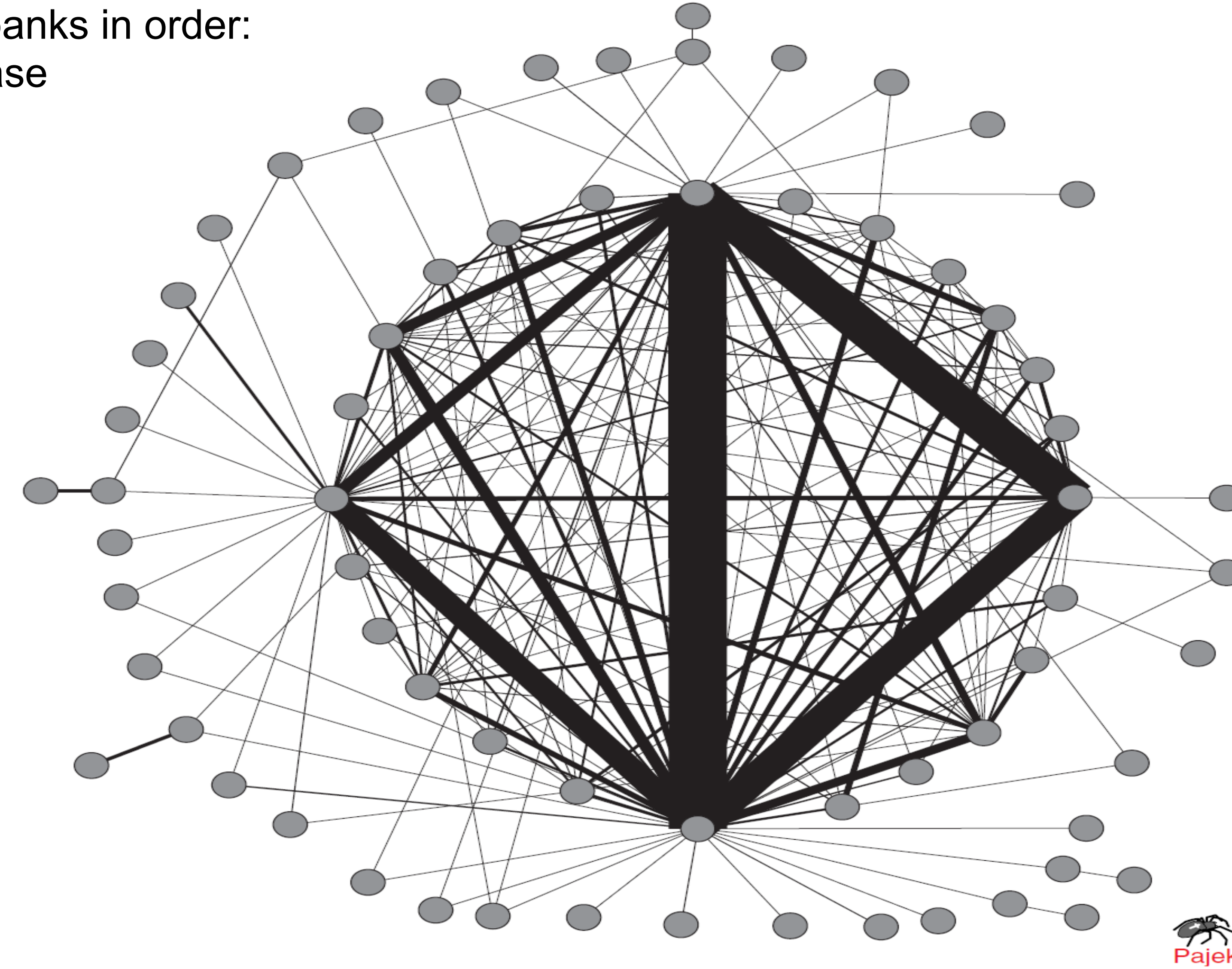
The Structure of Romantic and Sexual Relations at "Jefferson High School"



Military Alliances 2000

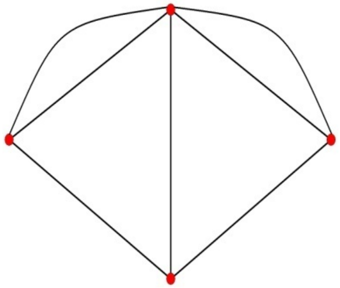


2012 largest banks in order:
JPMorganChase
B.of A.
Citigroup
W.fargo
Goldman
Metlife
MorganSt.



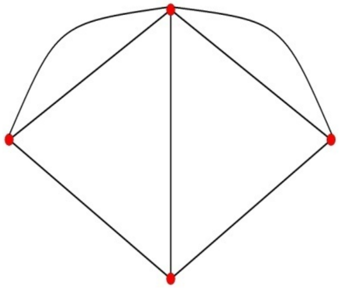
Fedwire Interbank payments, nodes accounting for 75% of total, Soromaki et al (2007), 25 nodes are completely connected

The Challenge:



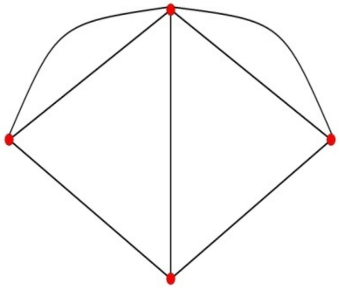
- How many networks on just 20 nodes?
- Person 1 could have 19 possible links, person 2 could have 18 not counting 1, ... total = 190
- So 190 possible links, each could either be present or not, so 2 x 2 x 2 ... 190 times = 2^{190} networks
- Atoms in the universe: somewhere between 2^{158} and 2^{246}

Simplifying the Complexity



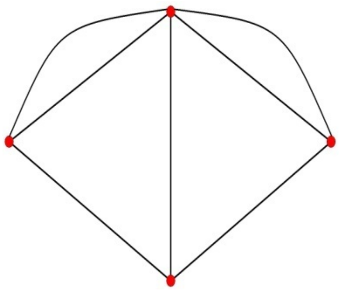
- Global patterns of networks
  - path lengths
  - degree distributions...
- Segregation Patterns: node types and homophily
- Local Patterns
  - Clustering
  - Support…
- Positions in networks
  - neighborhoods
  - Centrality, influence...

Simplifying the Complexity



- Global patterns of networks
  - path lengths
  - degree distributions...
- Segregation Patterns: node types and homophily
- Local Patterns
  - Clustering
  - Support...
- Positions in networks
  - neighborhoods
  - Centrality, influence...

Representing Networks



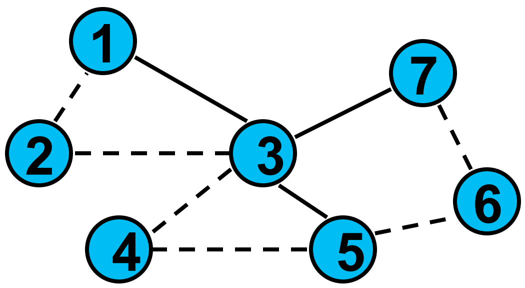
- $N=\{1,\ldots,n\}$ nodes, vertices, players

- $g \in \{0,1\}^{n \times n}$ (or $g \in [0,1]^{n \times n}$) relationships

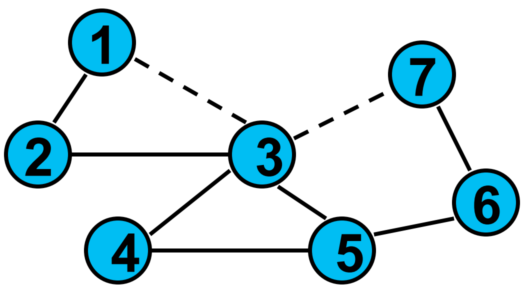
- $g_{ij} > 0$ indicates a link or edge between i and j

- Network (N,g)

Paths, Geodesics...

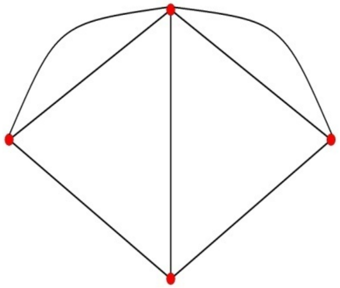


Path from 1 to 7



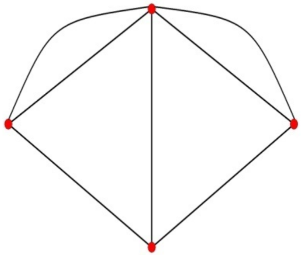
Geodesic: shortest Path from 1 to 7

Measures:



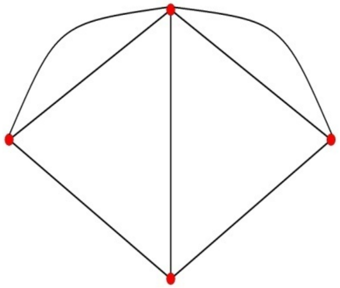
- Diameter – largest geodesic
  - if unconnected, of largest component...
- Average path length (less prone to outliers)
- *Affects speed of diffusion, contagion...*

Small average path length and diameter

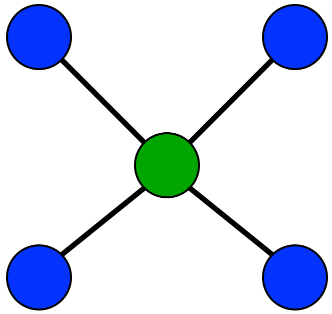


- Milgram (1967) letter experiments
  - median 5 for the 25% that made it
- Co-Authorship studies
  - Grossman (1999) Math mean 7.6, max 27,
  - Newman (2001) Physics mean 5.9, max 20
  - Goyal et al (2004) Economics mean 9.5, max 29
- Facebook
  - Ugander et al (2011) – mean 4.7 (99.9% of 720M pages)

Intuition Small Distances:

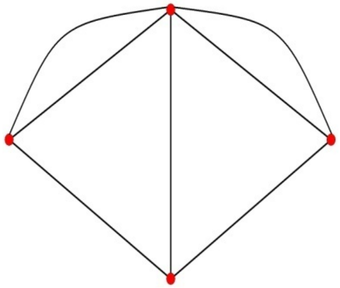


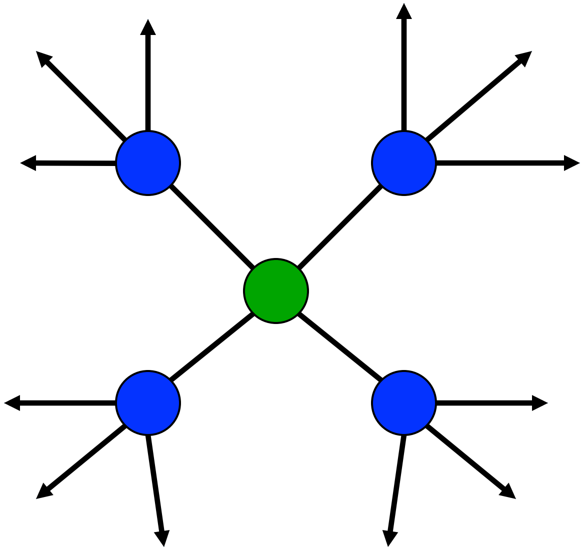
1 step: Reach d nodes,



Easy Calculation - Cayley Tree:
each node besides leaves has d links

Ideas:

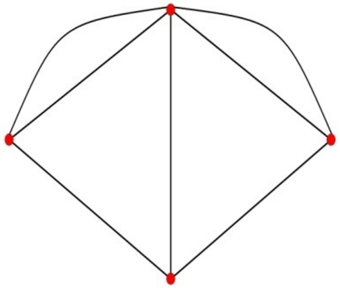


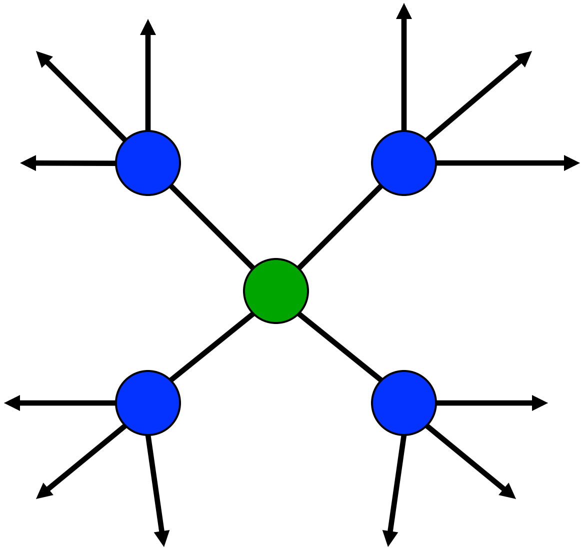


1 step: Reach d nodes,

then d(d-1),

Ideas:

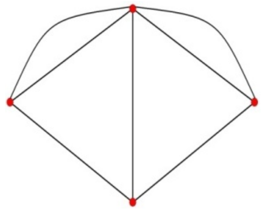




1 step: Reach d nodes,

then d(d-1),

then $d(d-1)^2$,



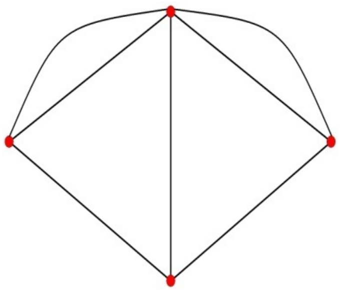
- Moving out l links from root in each direction reaches $d + d(d-1) + \dots\ d(d-1)^{l-1}$ nodes

- This is $d((d-1)^{l} - 1)/(d-2)$ nodes or roughly $(d-1)^{l}$

- To reach n-1, need roughly $(d-1)^{l} = n-1$ or

- $l = \log(n-1)/\log(d-1) \sim \log(n)/\log(d)$

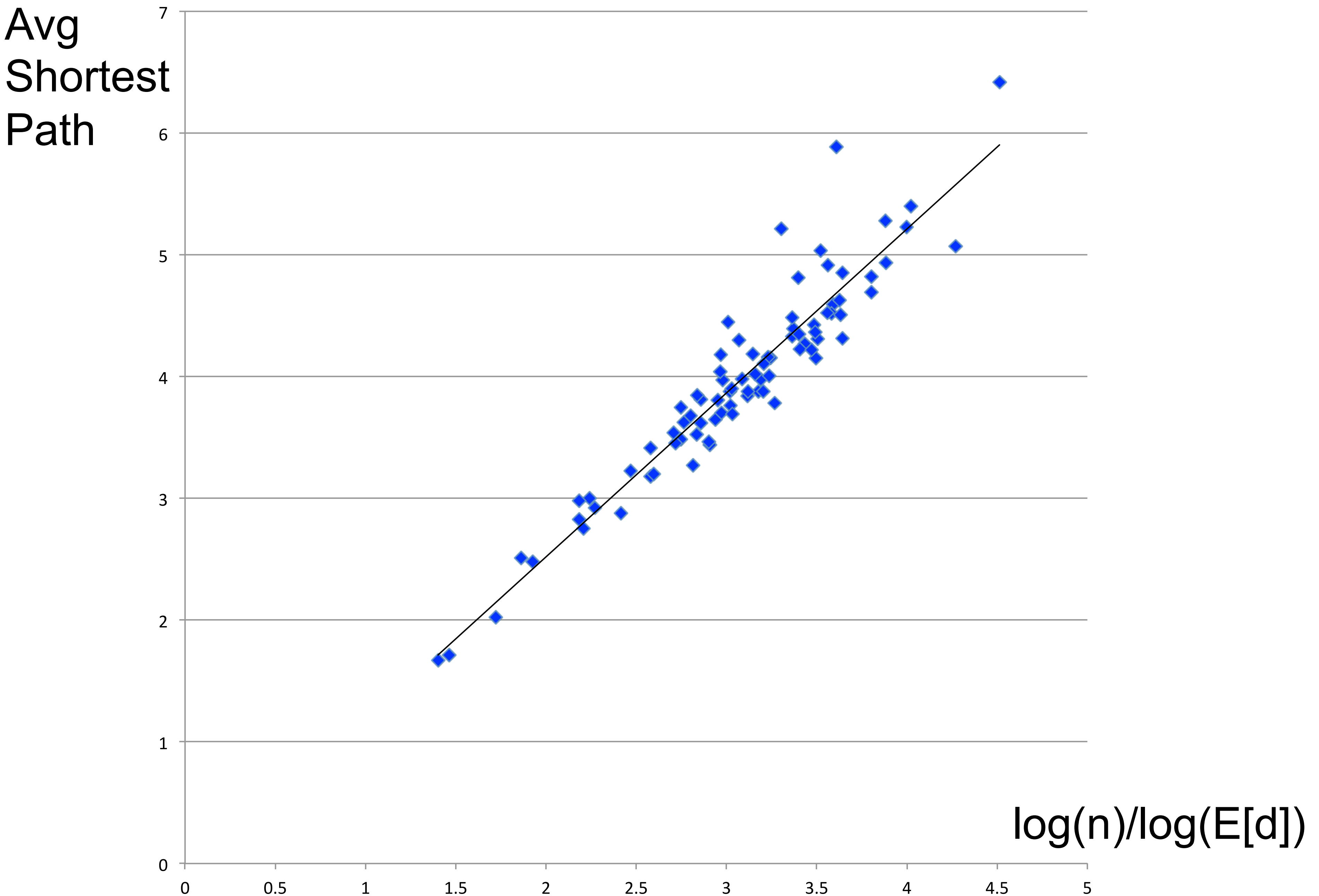
Same for Random Graphs



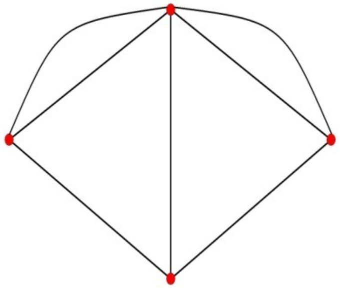
Theorem(s): For many classes of large random graphs average distance and max distance are proportional to log(n)/ log (E(d)).

[Ergos-Renyi (1959, 1960), Chung-Lu (2002), Jackson (2008b)]

84 High Schools – Ad Health

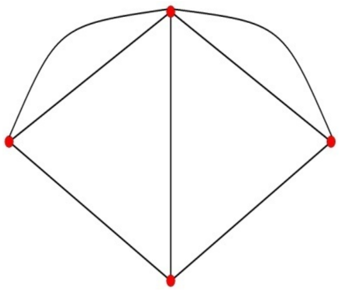


Small Worlds/Six Degrees of Separation



- n = 6.7 billion (world population)
- d = 50 (friends, relatives...)
- log(n)/log(d) is about 6 !!

Neighborhood and Degree



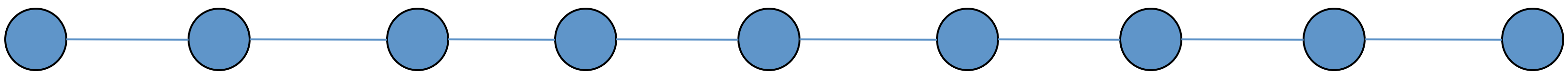
- Neighborhood: $N_i(g) = \{ j \mid ij \text{ in } g \}$ (convention ii not in g)

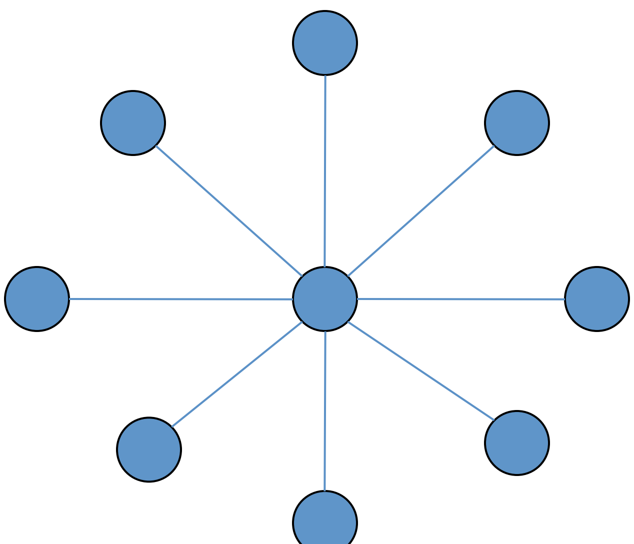
- Degree: $d_i(g) = \# N_i(g)$

- How is Degree distributed in a network?

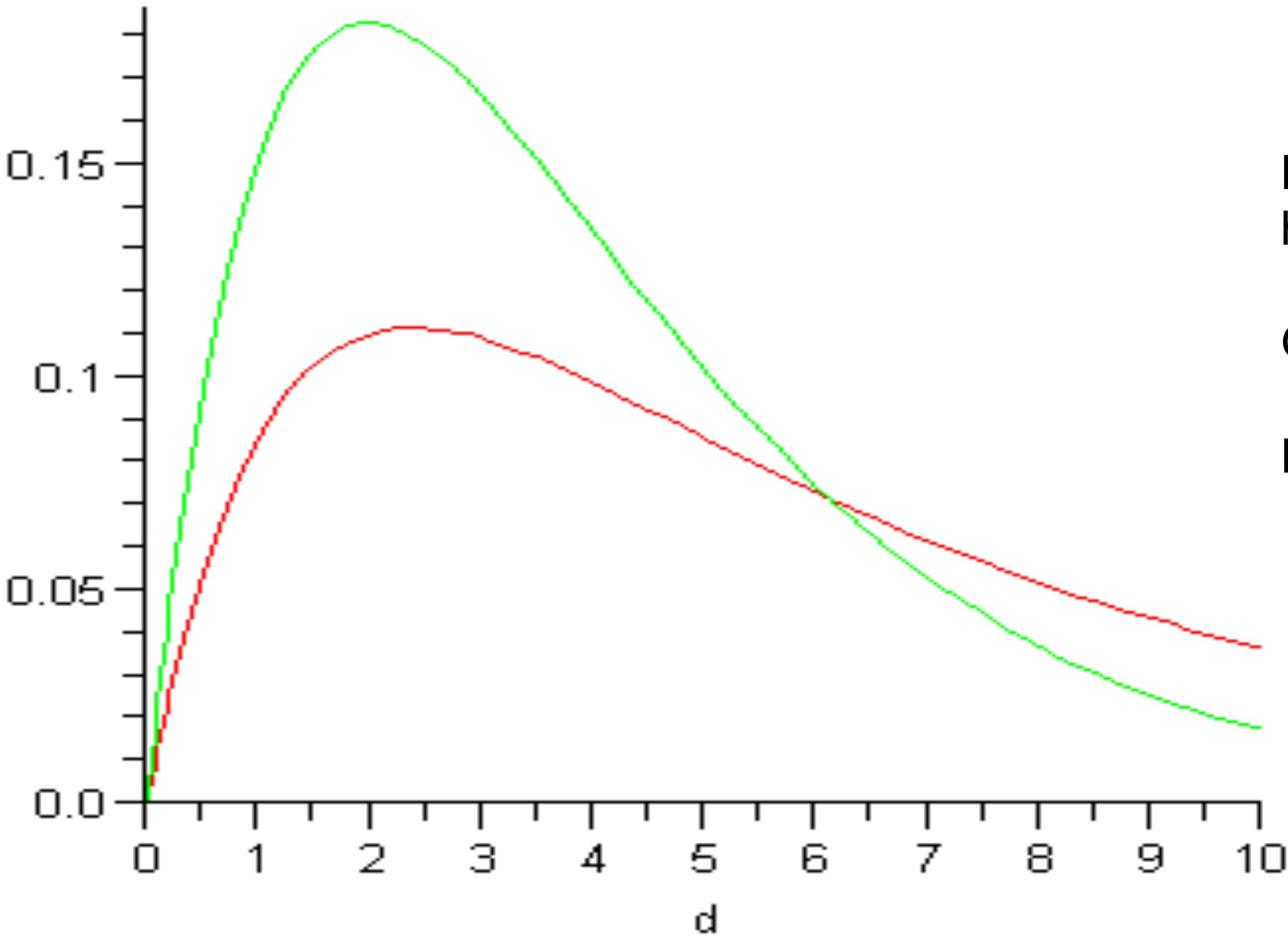
Degree Distributions

- Average degree tells only part of the story:





Coauthor versus Romance

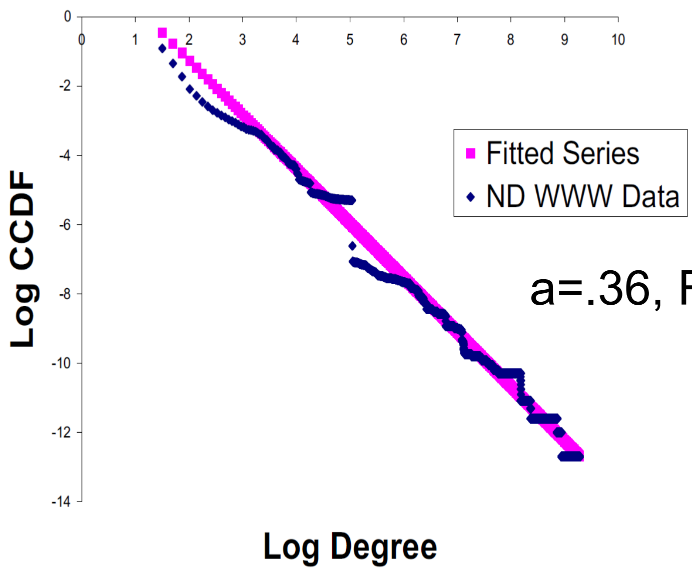


Prob given neighbor has degree

Green – romance

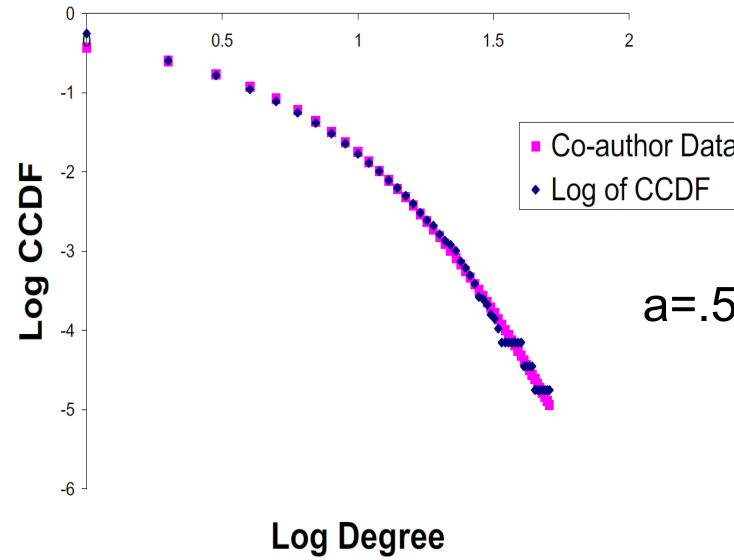
Red - coauthor

Fitting WWW Data



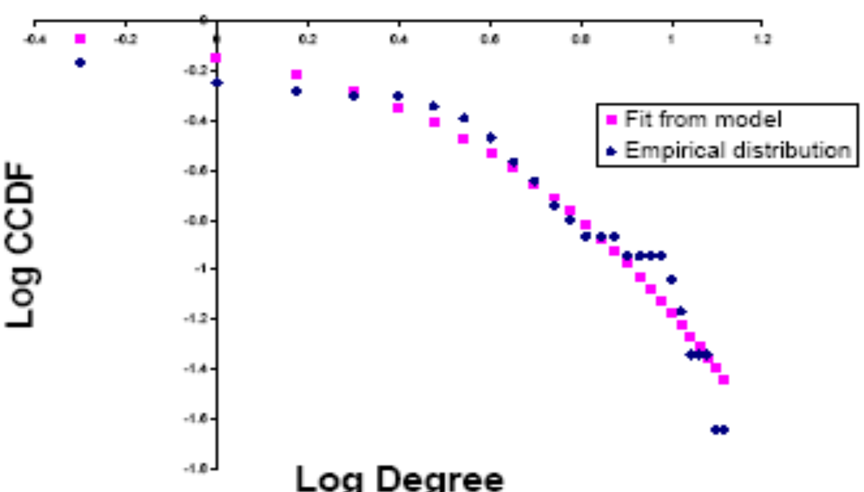
a=.36, R^2=.97

Fitting Co-author Data



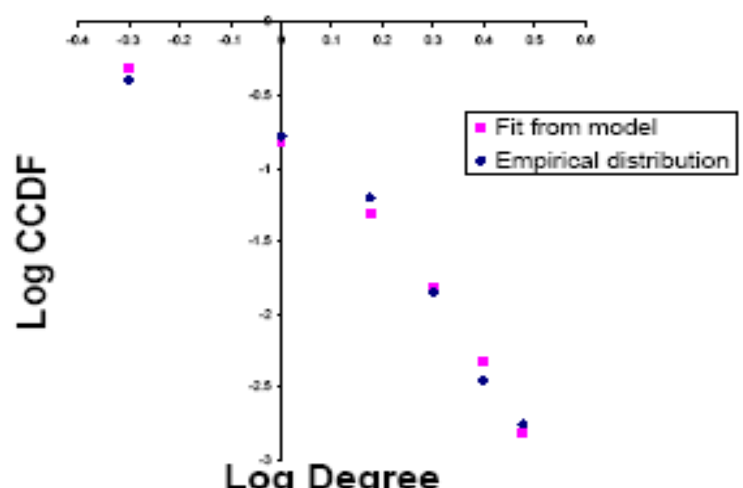
a=.56, R^2=.99

Ham Radio



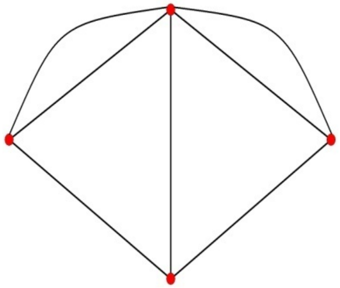
a=.59, R^2=.94

High School Romance



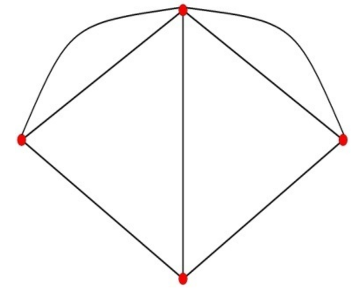
a=1, R^2=.99

Simplifying the Complexity



- Global patterns of networks
  – path lengths
  – degree distributions...
- Segregation Patterns: node types and homophily
- Local Patterns
  – Clustering
  – Support...
- Positions in networks
  – neighborhoods
  – Centrality, influence...

Homophily:



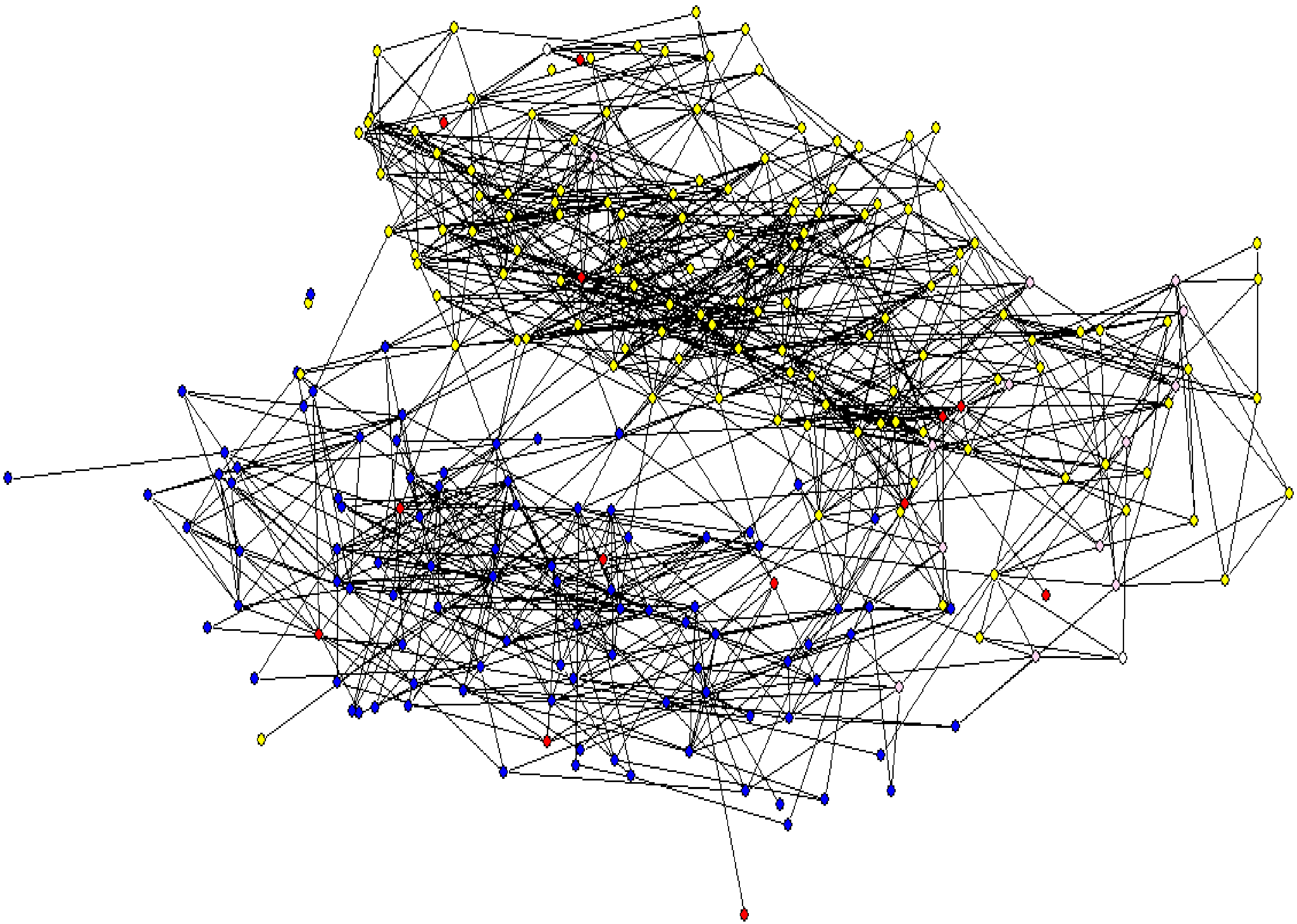
``Birds of a Feather Flock Together'' - Philemon Holland (1600 - ``As commonly birds of a feather will flye together'')

Effects: *heterogeneity of behavior, beliefs, culture; peer effects (endogeneity!!); poverty traps, ...*

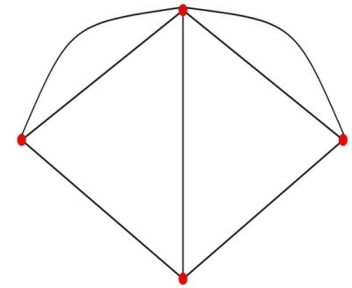
- age, race, gender, religion, profession....
  - Lazarsfeld and Merton (1954) ``Homophily''
  - Shrum et al (gender, ethnic, 1988…), Blau (professional 1974, 1977), Marsden (variety, 1987, 1988), Moody (grade, racial, 2001…), McPherson (variety,1991…)…

Blue: Blacks
Reds: Hispanics
Yellow: Whites
White: Other

Currarini, Jackson, Pin 09, 10



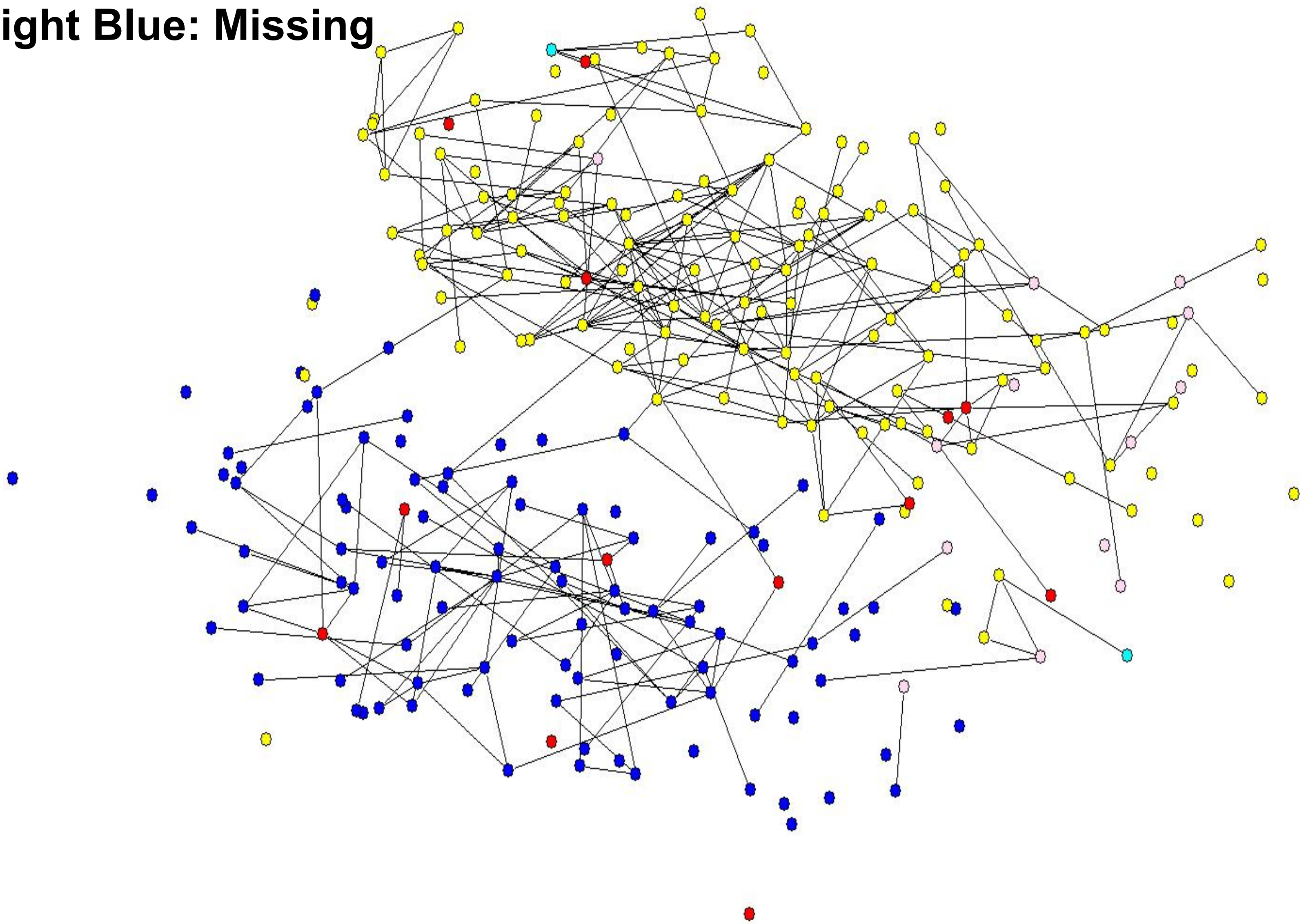
Adolescent Health, High School in US:

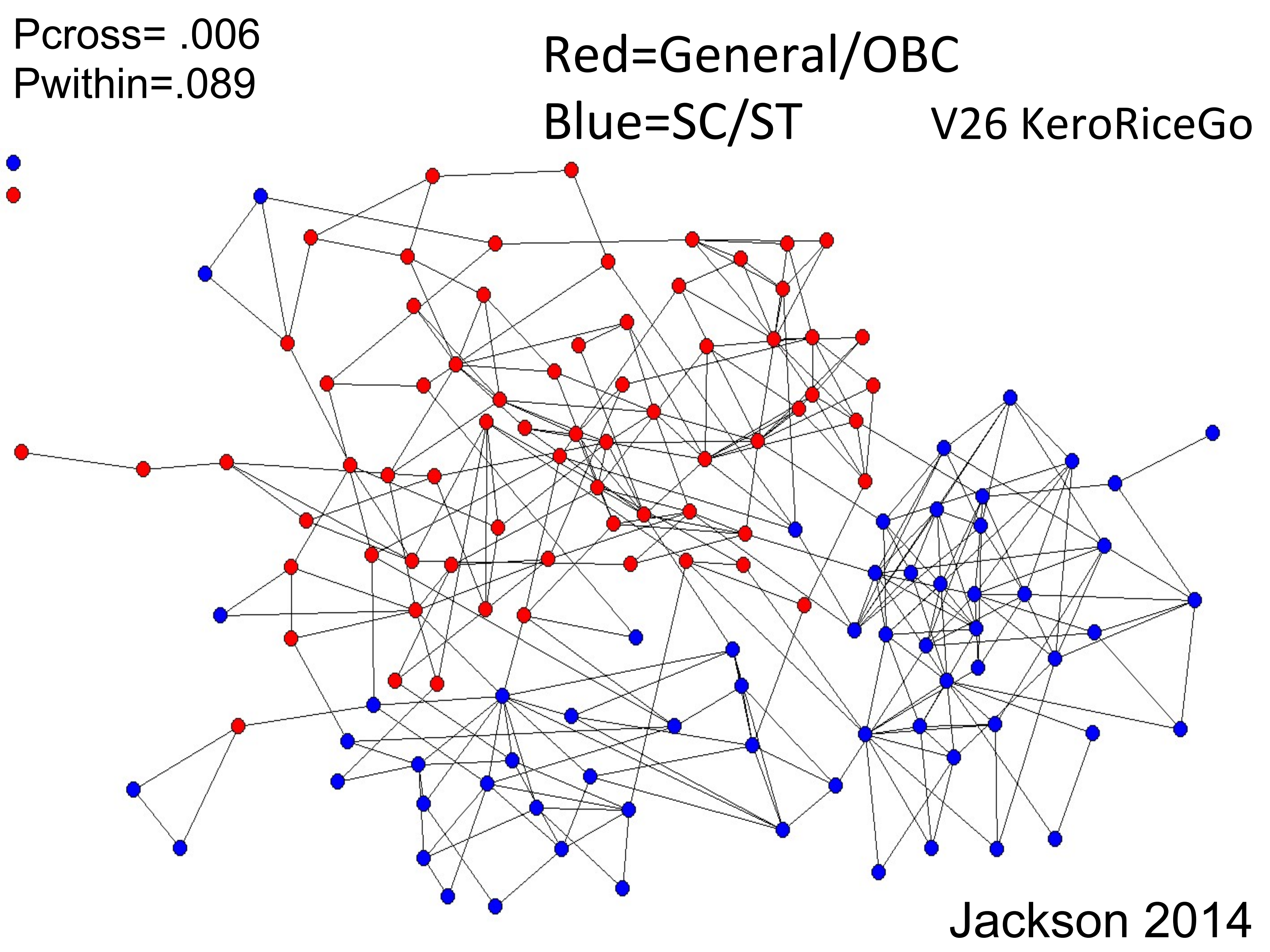


Percent:	52	38	5	5
	White	Black	Hispanic	Other
White	86	7	47	74
Black	4	85	46	13
Hispanic	4	6	2	4
Other	6	2	5	9
	100	100	100	100

Blue: Black
Reds: Hispanic
Yellow: White
Pink: Other
Light Blue: Missing

"strong friendships"
cross group links less than half as frequent
Jackson (07)

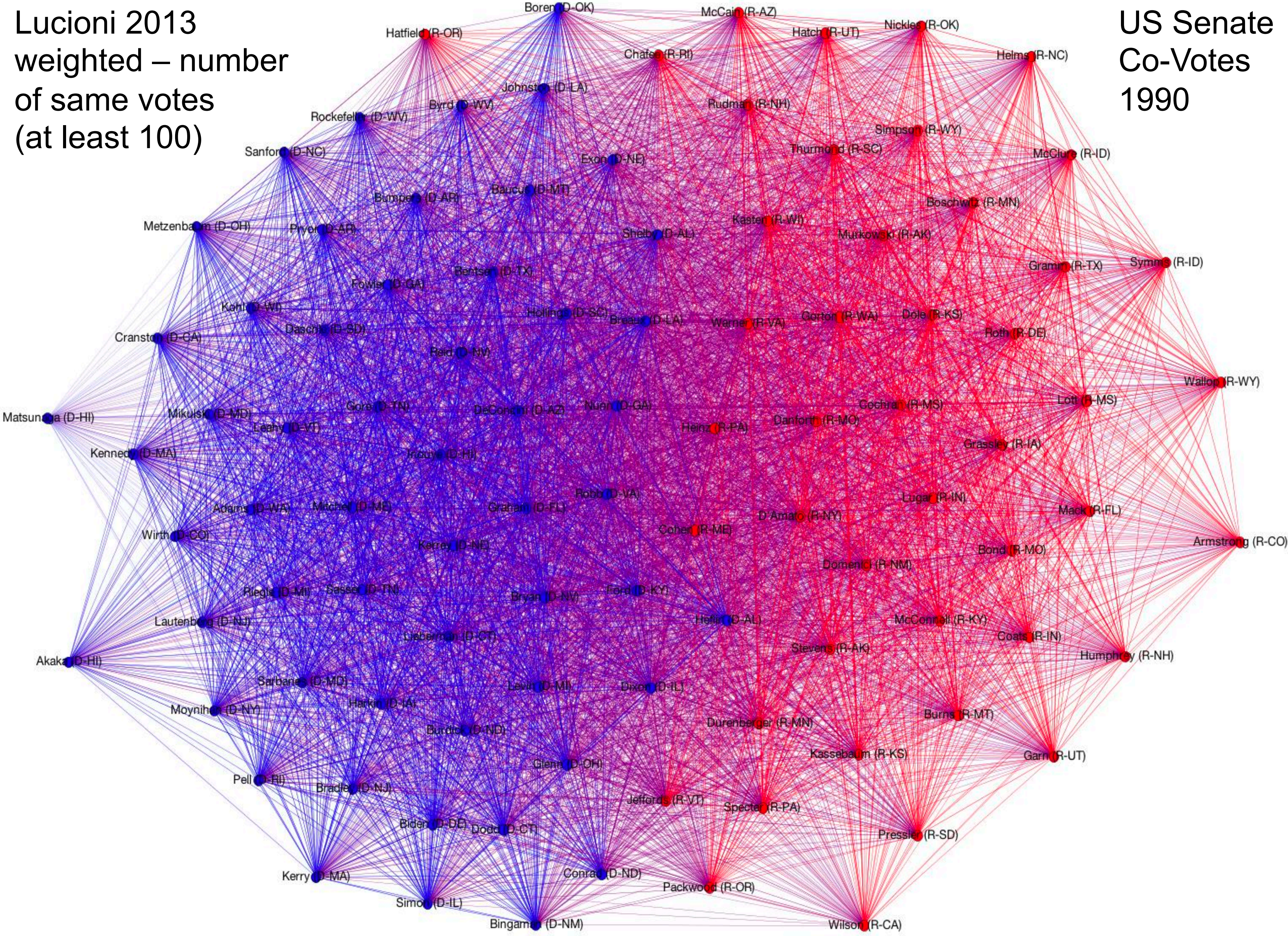




Pcross= .006
Pwithin=.089
Red=General/OBC
Blue=SC/ST
V26 KeroRiceGo
Jackson 2014

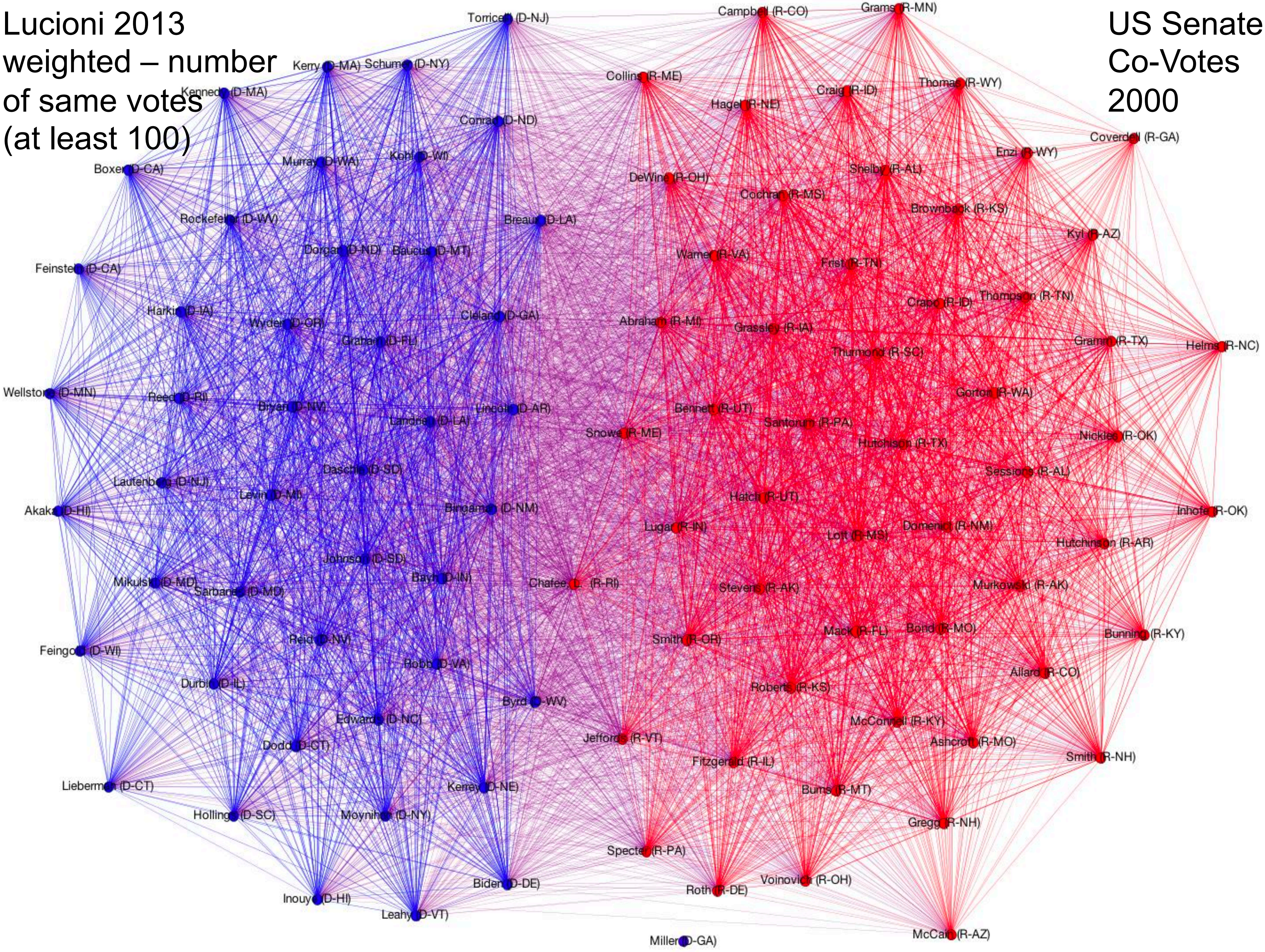
Lucioni 2013
weighted – number
of same votes
(at least 100)

US Senate
Co-Votes
1990



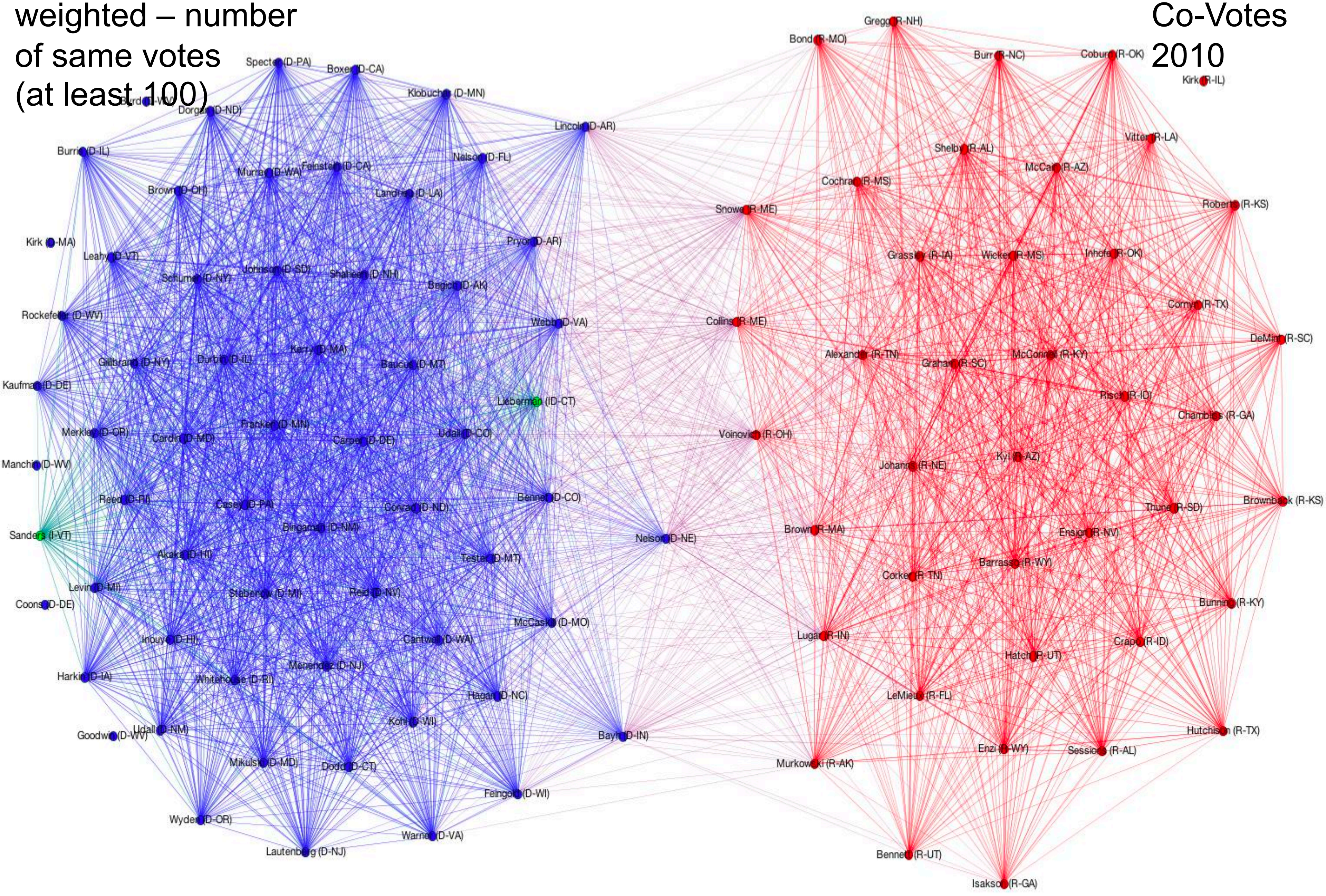
Lucioni 2013
weighted – number
of same votes
(at least 100)

US Senate
Co-Votes
2000

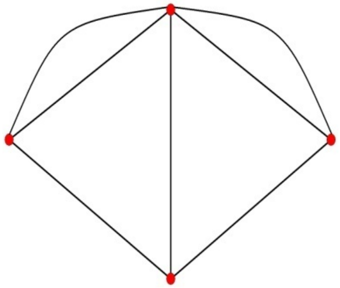


Lucioni 2013
weighted – number
of same votes
(at least 100)

US Senate
Co-Votes
2010

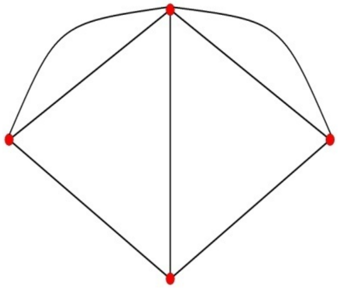


Simplifying the Complexity



- Global patterns of networks
  - path lengths
  - degree distributions...
- Segregation Patterns: node types and homophily
- Local Patterns
  - Clustering
  - Support...
- Positions in networks
  - neighborhoods
  - Centrality, influence...

Influence/Centrality/Power

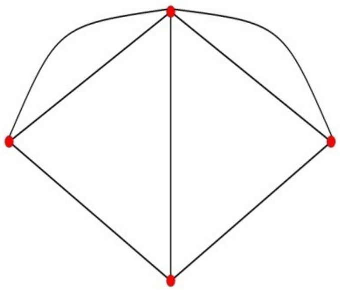


- Economists care about networks because of externalities: interactions between nodes

- Heterogeneity of nodes' influence not just due to characteristics, but also due to network position

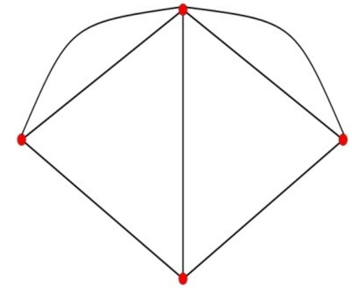
- How to capture this?   Depends on nature of interaction...

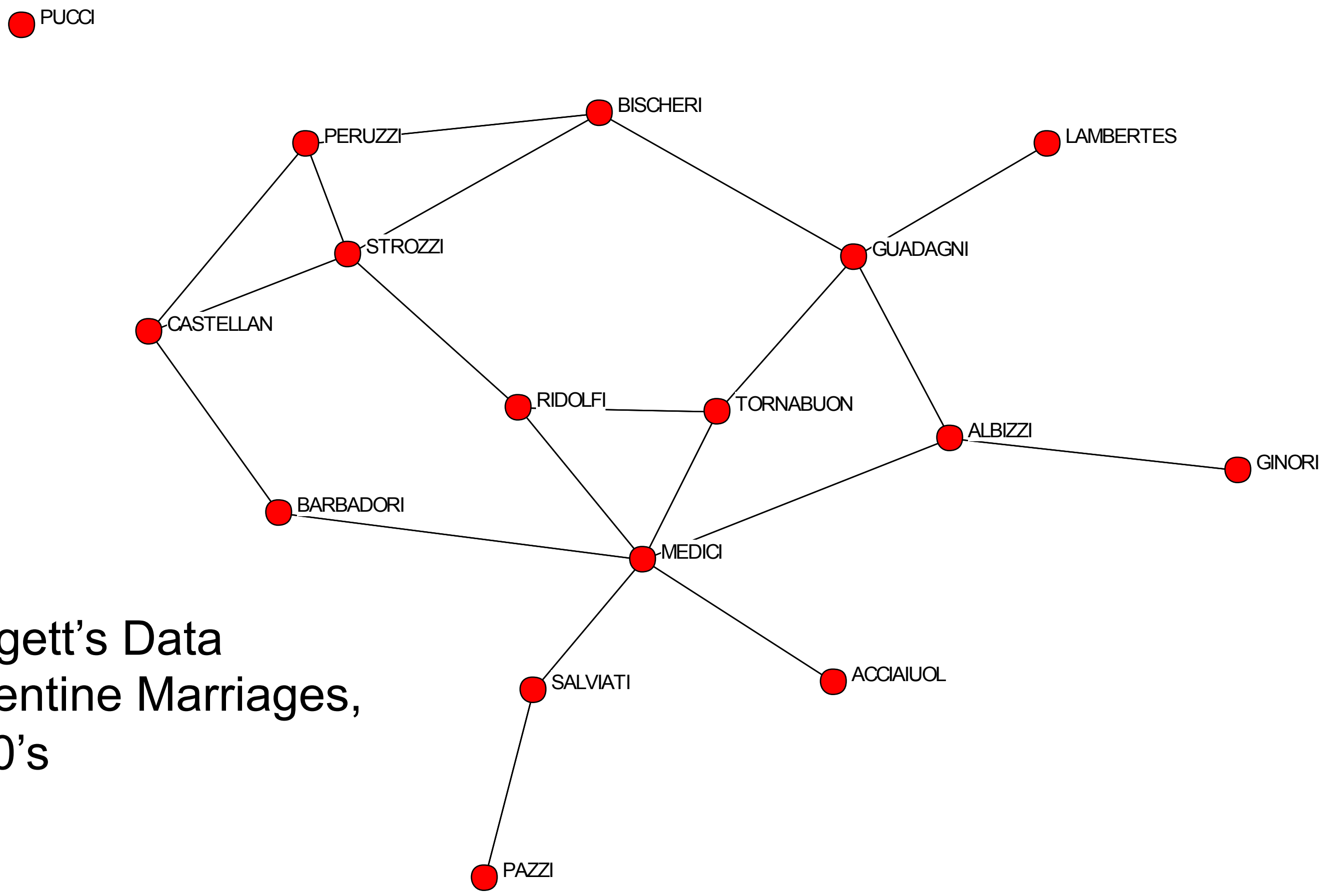
Degree Centrality



- How ``connected'' is a node?
  - degree captures connectedness
  - normalize by n-1 - most possible

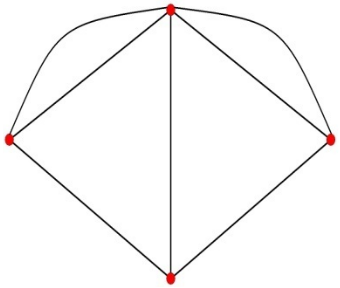
Degree Centrality





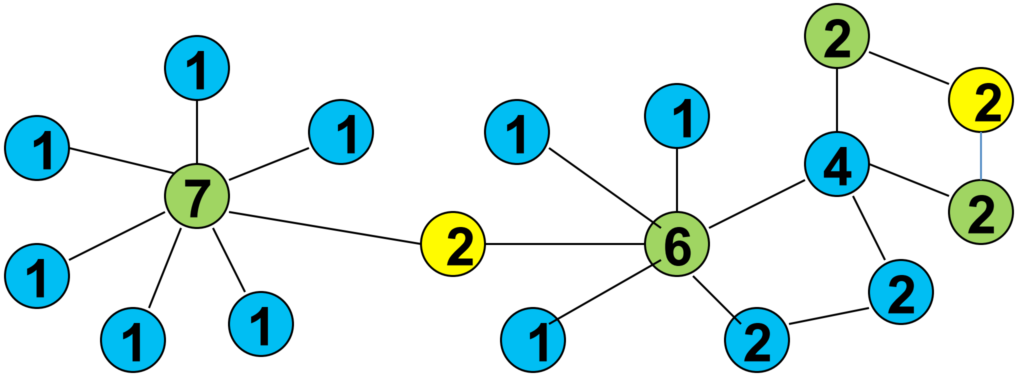
Medici = 6
Strozzi = 4
Guadagni = 4

Padgett's Data
Florentine Marriages,
1430's

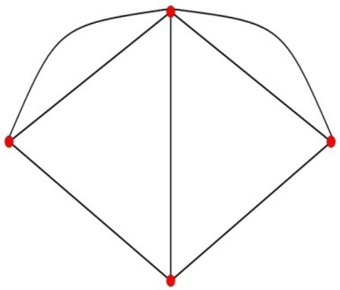


Degree Centrality?

- More reach if connected to a 6 and 7 than a 2 and 2?



Another Centrality

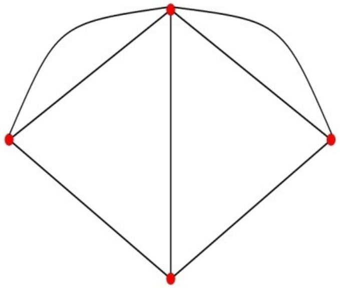


- Centrality proportional to the sum of neighbors' centralities

C_i proportional to $\sum_{j:\text{ friend of } i} C_j$

More connections matter, but also accounts for how central they are!

``Eigenvector Centrality''

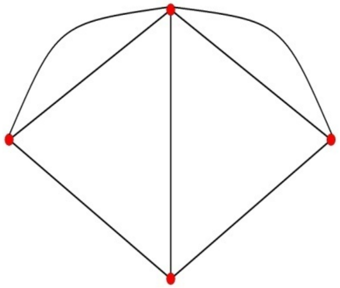


- Centrality is proportional to the sum of neighbors' centralities

C_i proportional to $\sum_{j:\text{ friend of } i} C_j$

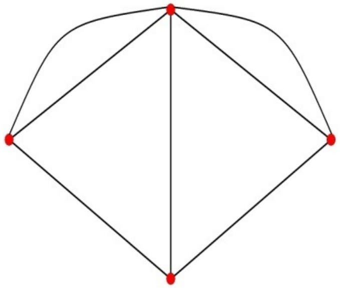
$$aC_i = \sum_j g_{ij} C_j$$

Centrality

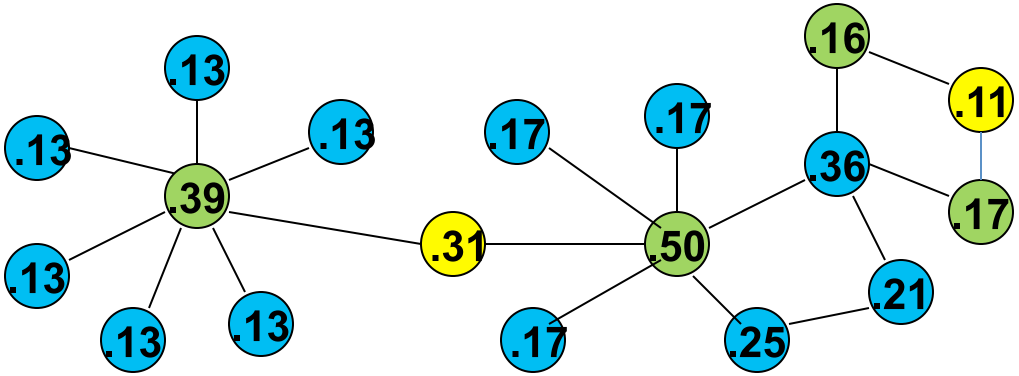


- Concepts related to eigenvector centrality:
- Google page rank: score of a page is proportional to the sum of the scores of pages linked to it
- Random surfer model: start at some page on the web, randomly pick a link, follow it, repeat...

Eigenvector Centrality

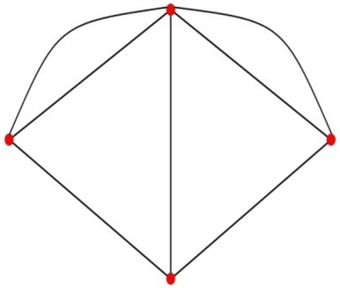


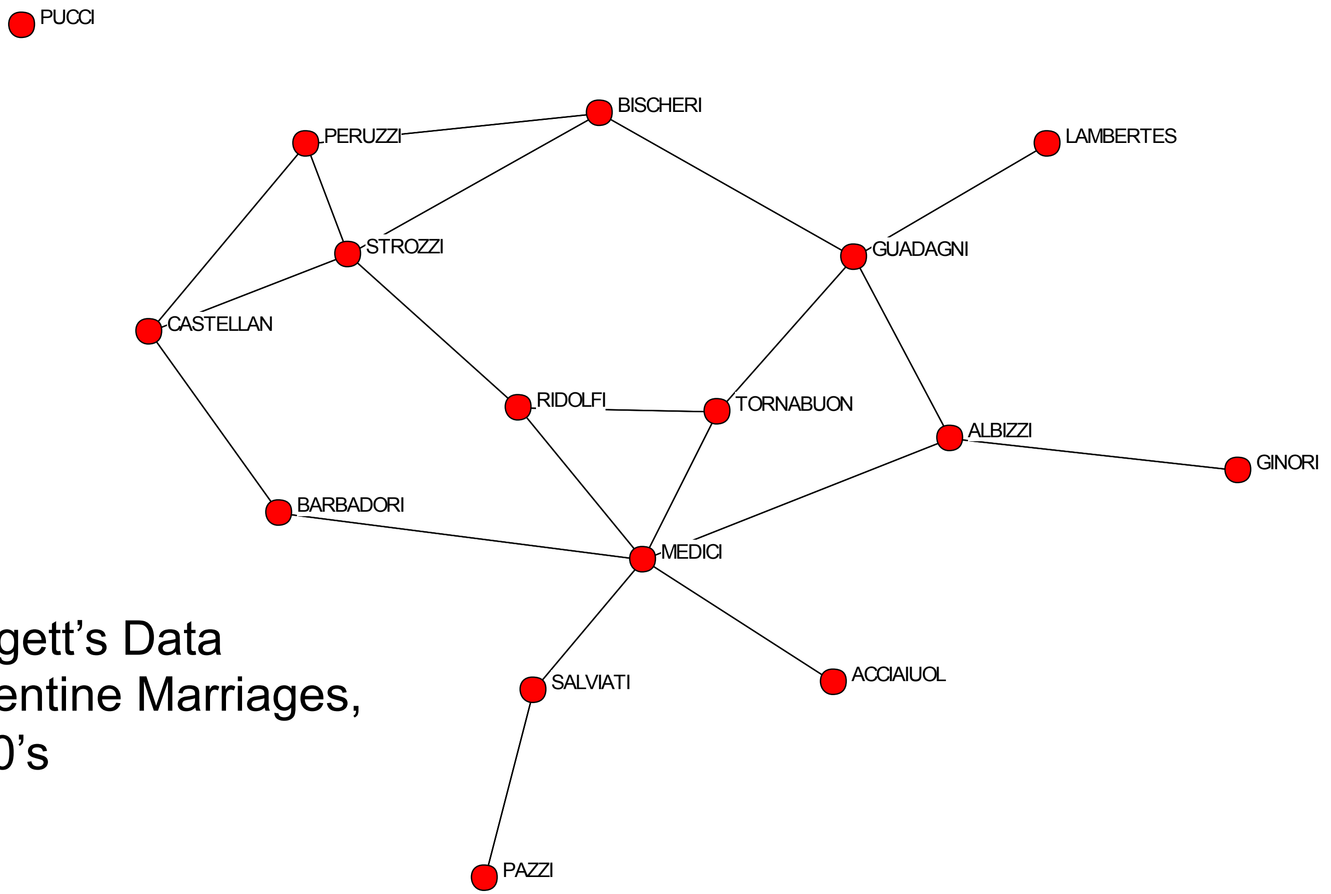
Now distinguishes more ``influential'' nodes



(eval 2.9 so prop to 1/2.9 C neighbors)

Eigenvector Centrality

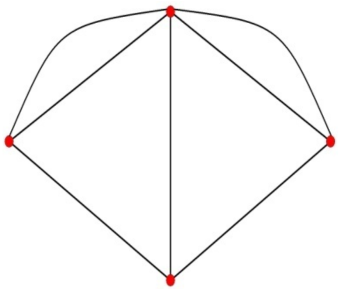




Medici = .430
Strozzi = .356
Guadagni = .289
Ridolfi=.341
Tornabuon=.326

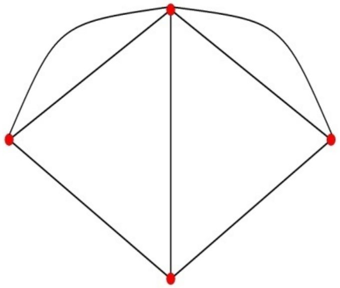
Padgett's Data
Florentine Marriages, 1430's

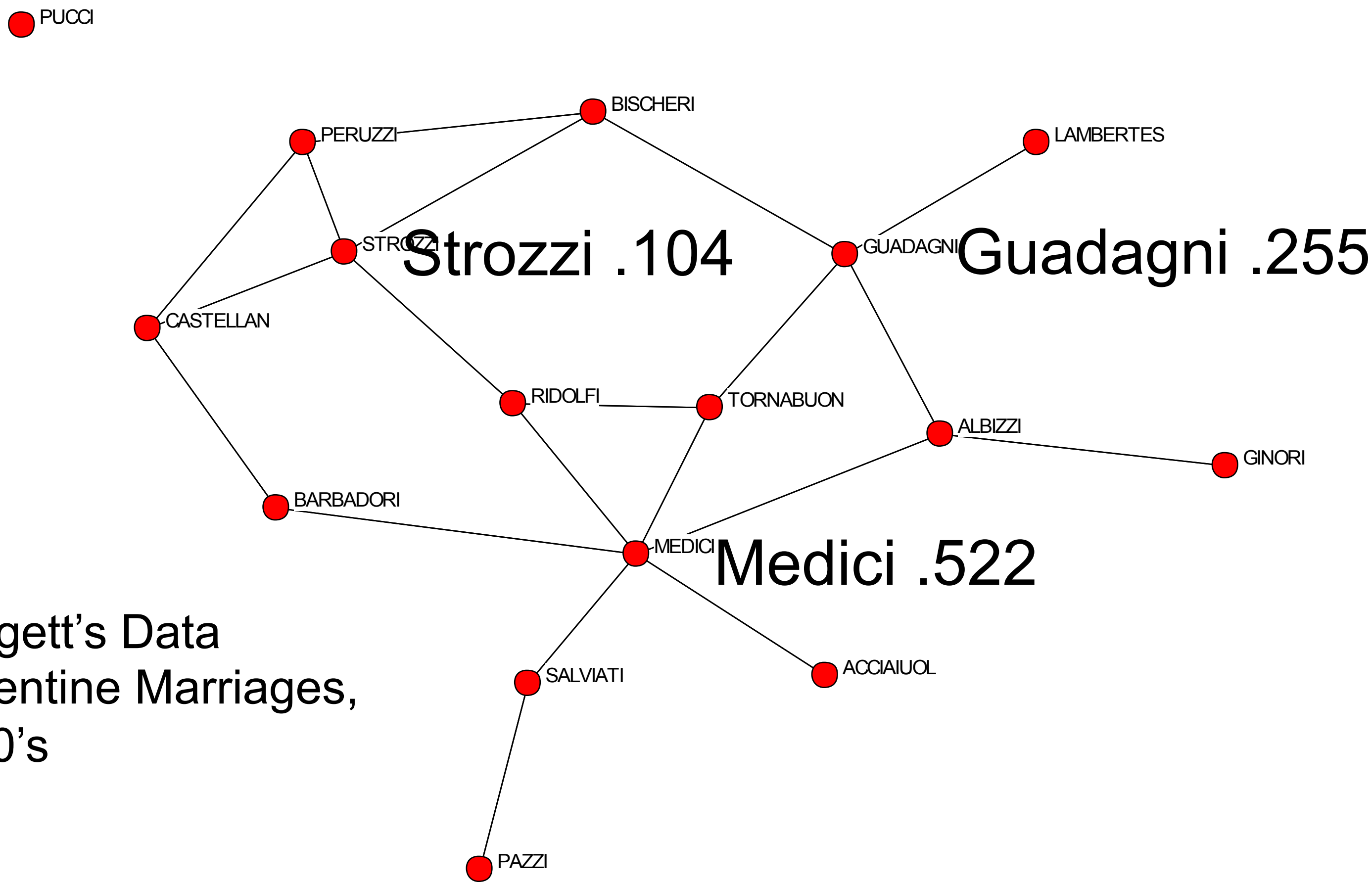
Betweenness (Freeman) Centrality



- P(i,j) number of geodesics between i and j
- $P_k(i,j)$ number of geodesics between i and j that k lies on
- $\sum_{i,j\neq k} [P_k(i,j)/ P(i,j)] / [(n-1)(n-2)/2]$

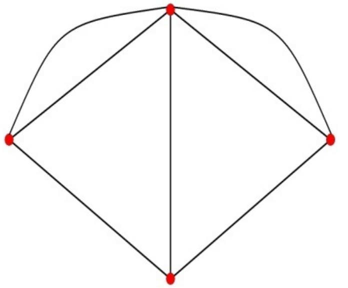
Betweenness Centrality





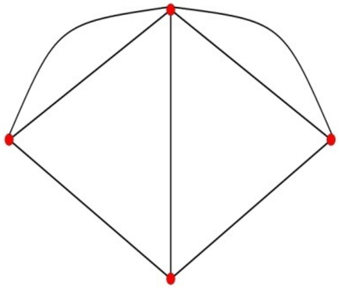
Padgett's Data
Florentine Marriages,
1430's

Centrality,
Four different things to measure:



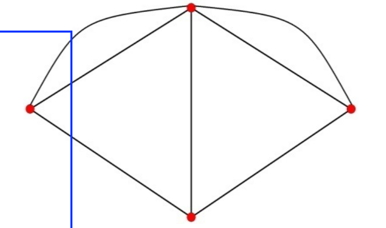
- Degree – connectedness
- Influence, Prestige, Eigenvectors – ``not what you know, but who you know..”
- Betweenness – importance as an intermediary, connector
- Closeness, Decay – ease of reaching other nodes

Application: Centrality affects diffusion



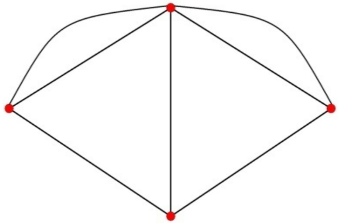
- Injection points:
  - How does centrality of first `infected/ informed' correlate with eventual diffusion?
  - Which centrality measures are predictive?

BCDJ 2013, Diffusion of Microfinance



- 75 rural villages in Karnataka, relatively isolated from microfinance initially
- BSS entered 43 of them and offered microfinance
- We surveyed villages before entry, observed network structure and various demographics
- Tracked microfinance participation over time

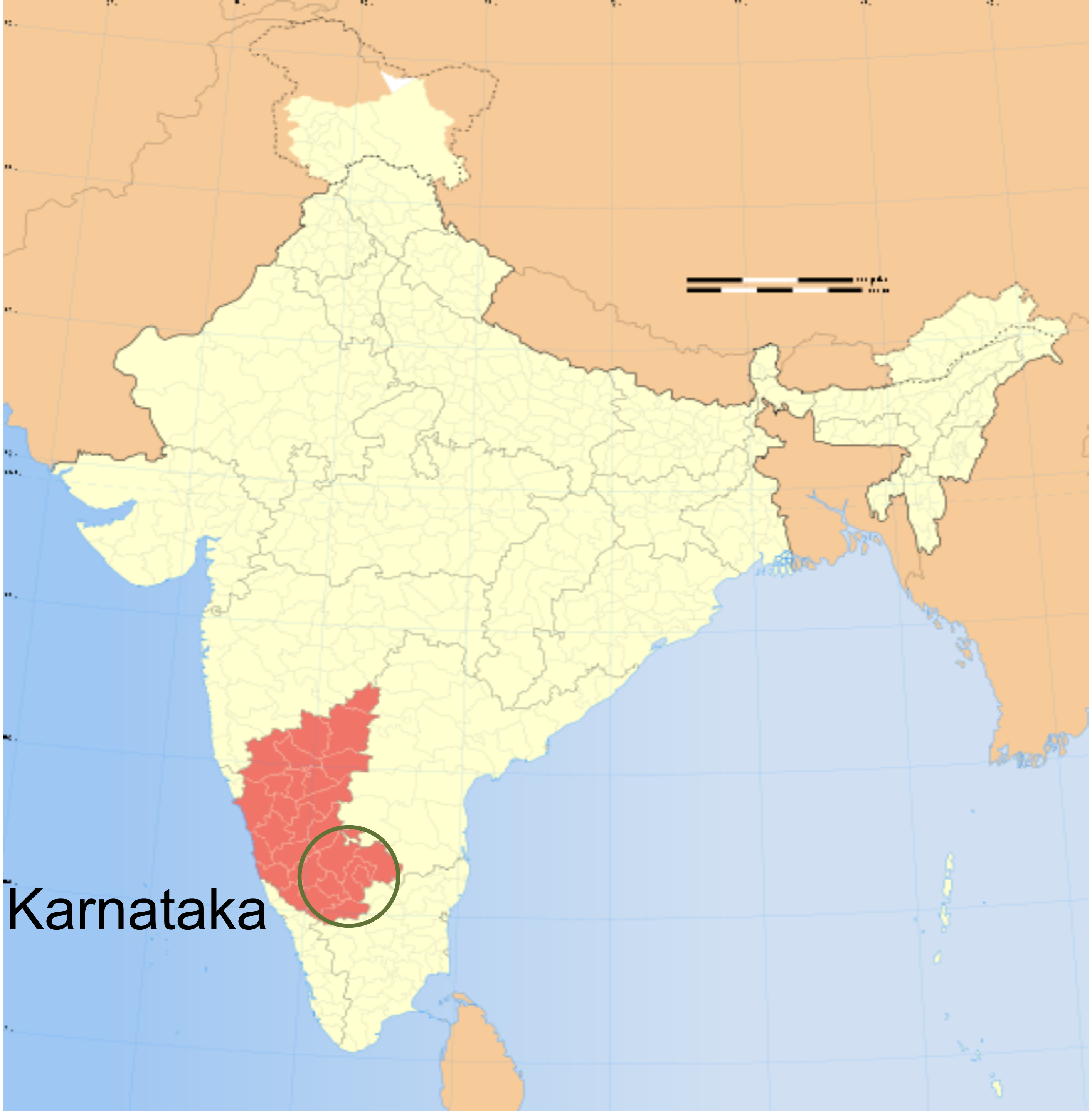
Background



- Microfinance participation varies widely even in otherwise similar villages
  - Near 0 in some places (7% min in our data)
  - Near 1/2 in others (44% max in our data)

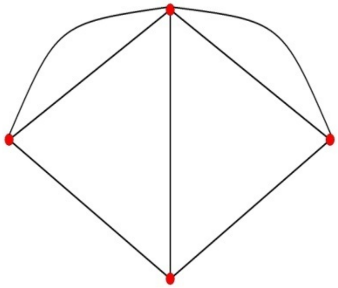
- Why?

- Word of mouth is essential in getting news out – How does it work/not?



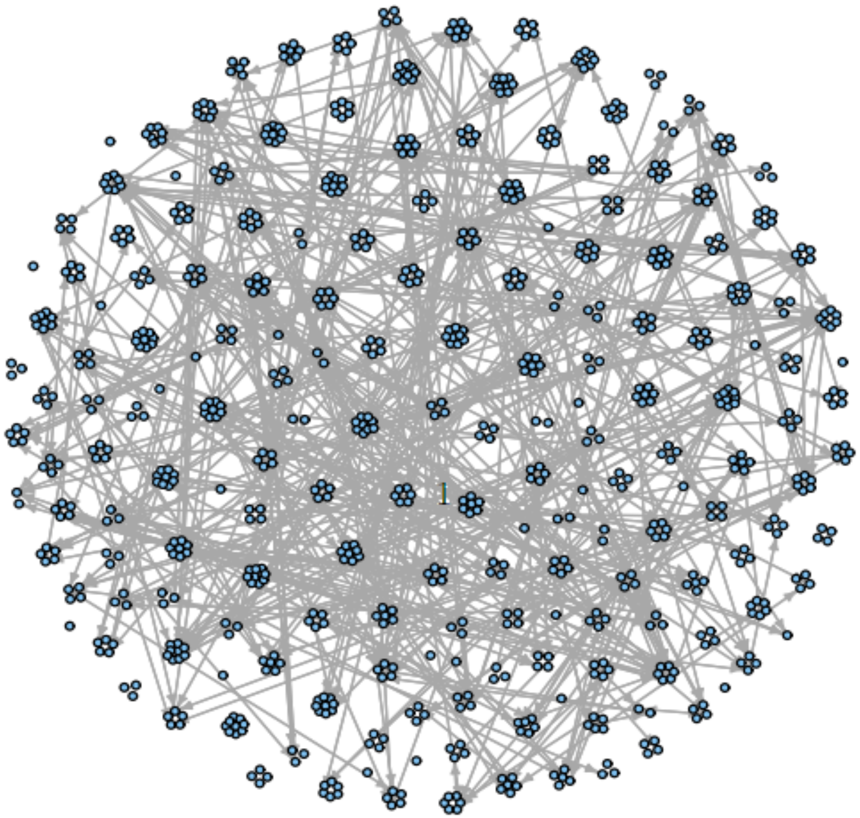
Karnataka

Background: 75 Indian Villages – Networks

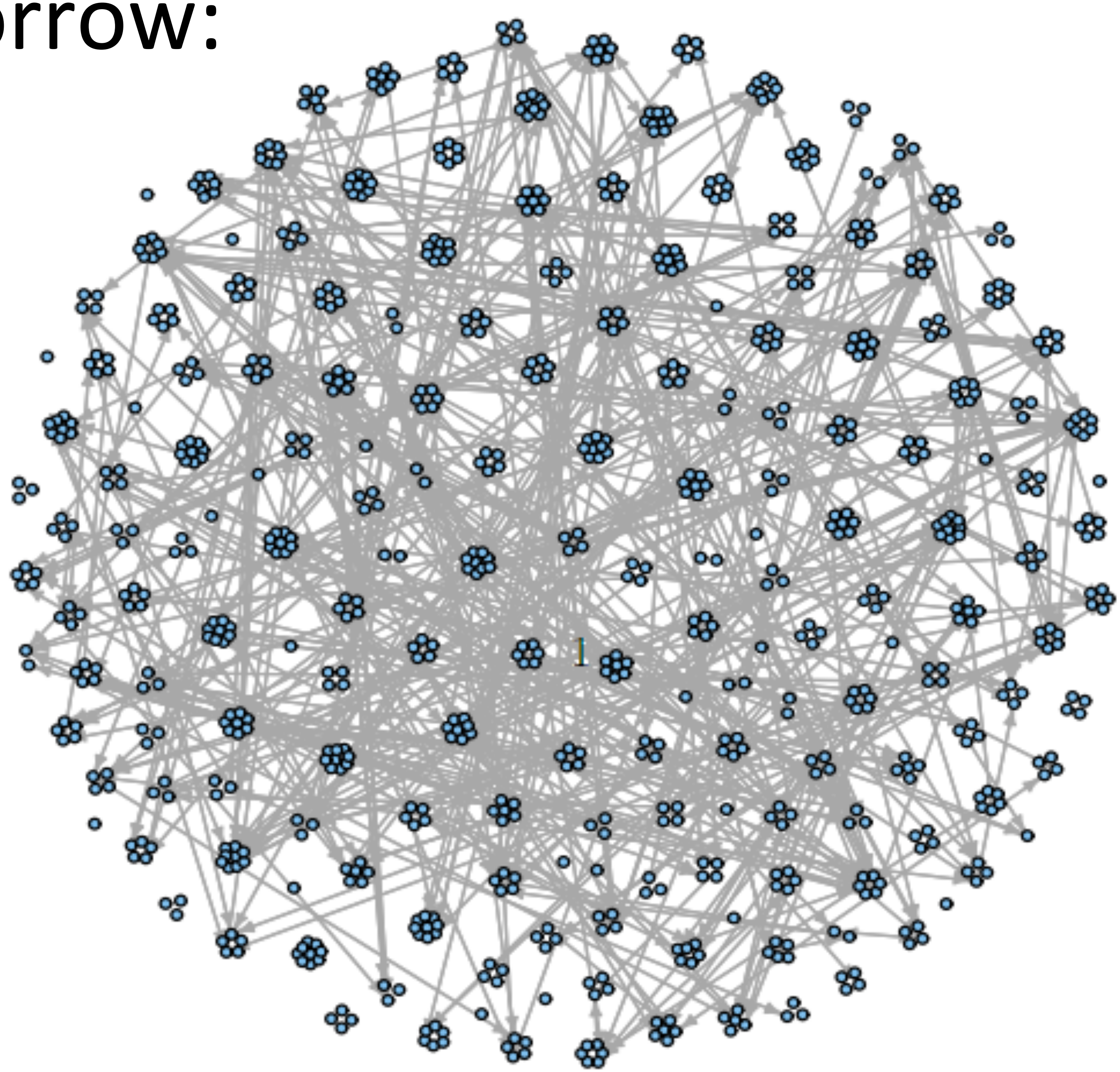


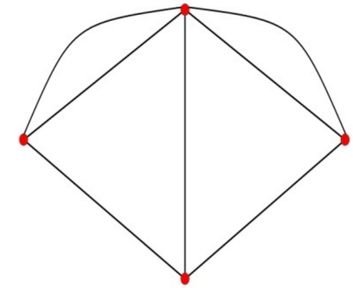
- ``Favor'' Networks:
  - both borrow and lend money
  - both borrow and lend kero-rice
- ``Social'' Networks:
  - both visit come and go
  - friends (talk together most)
- Others (temple, medical help…)

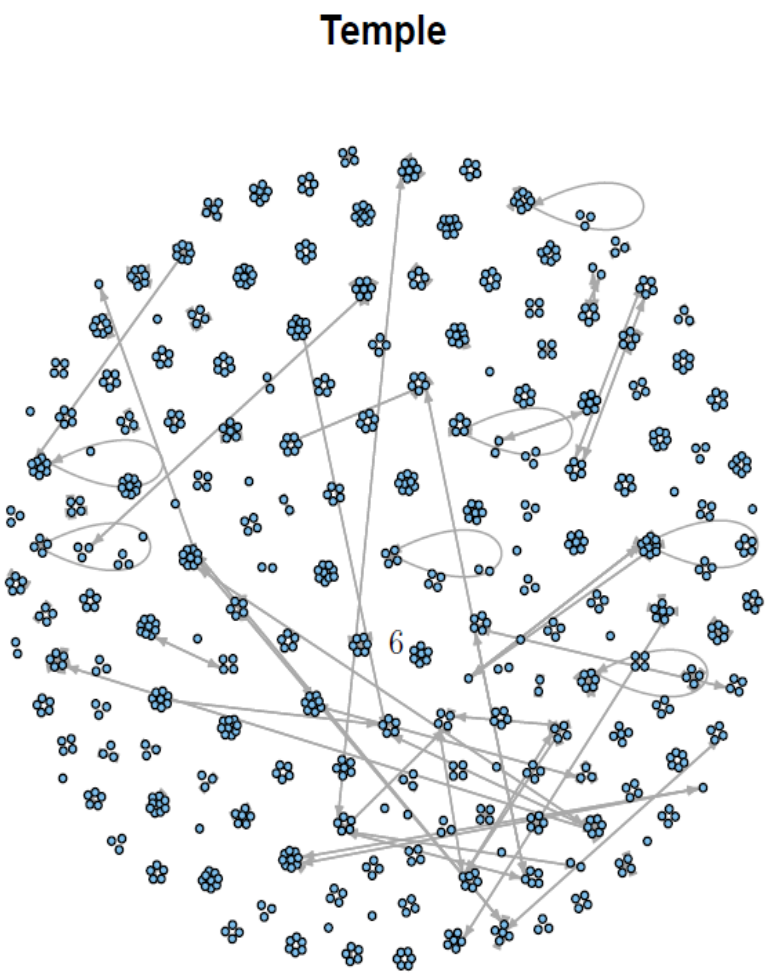
Borrow:

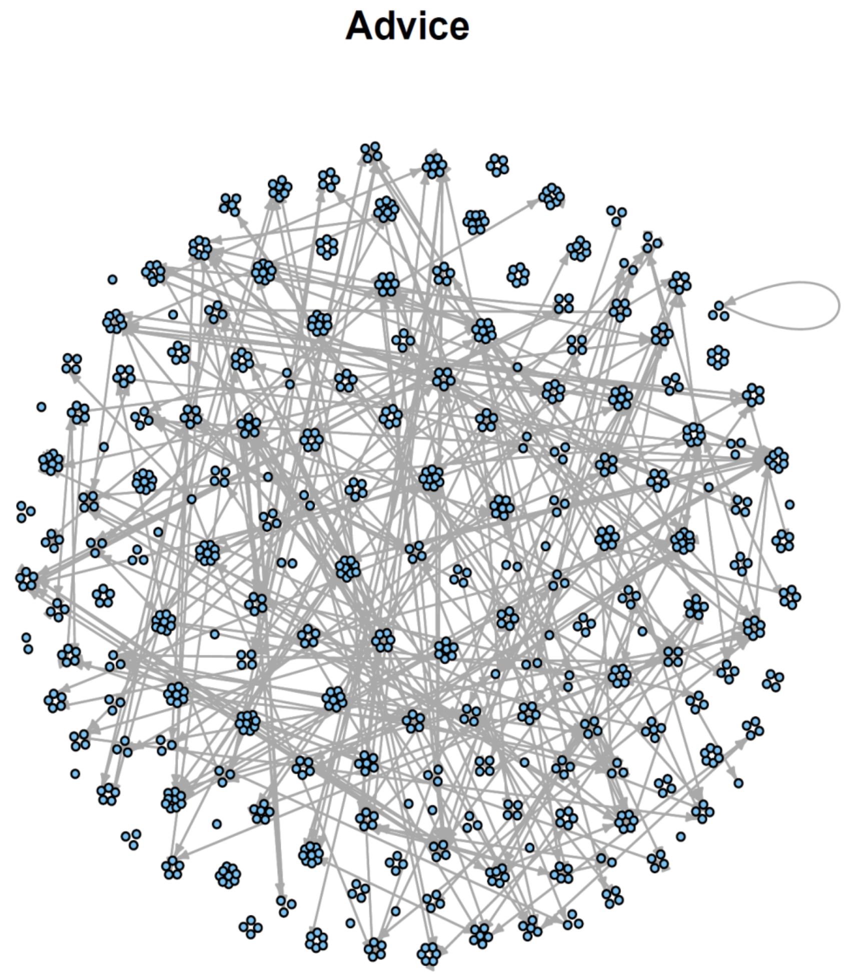


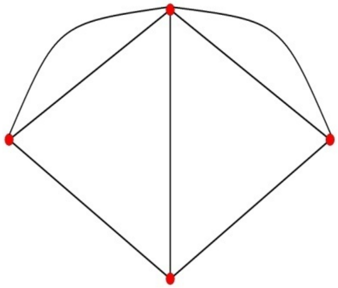
Borrow:



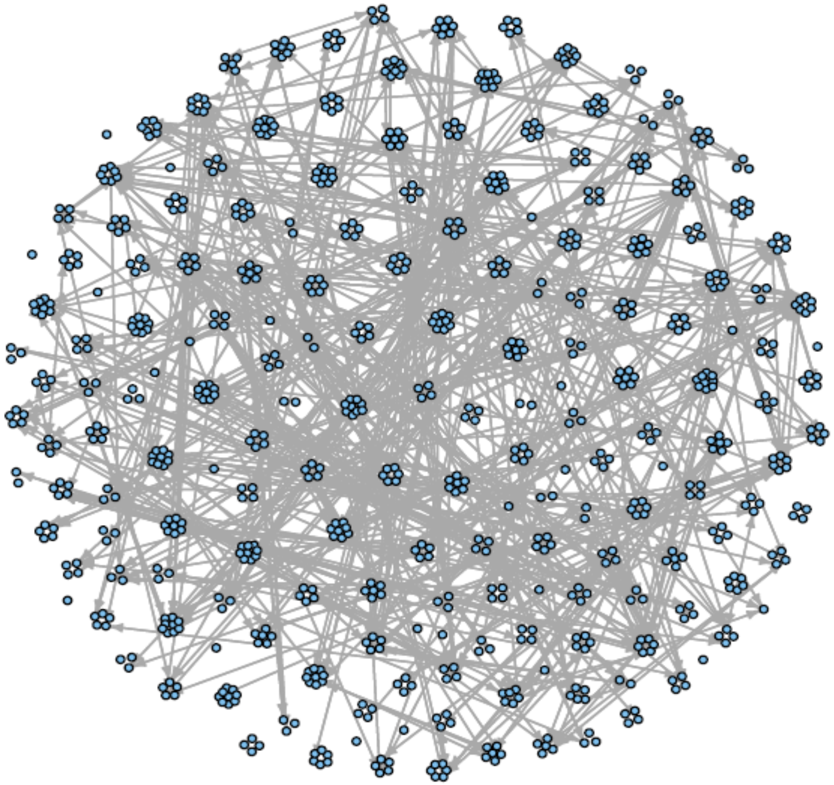



Temple
6

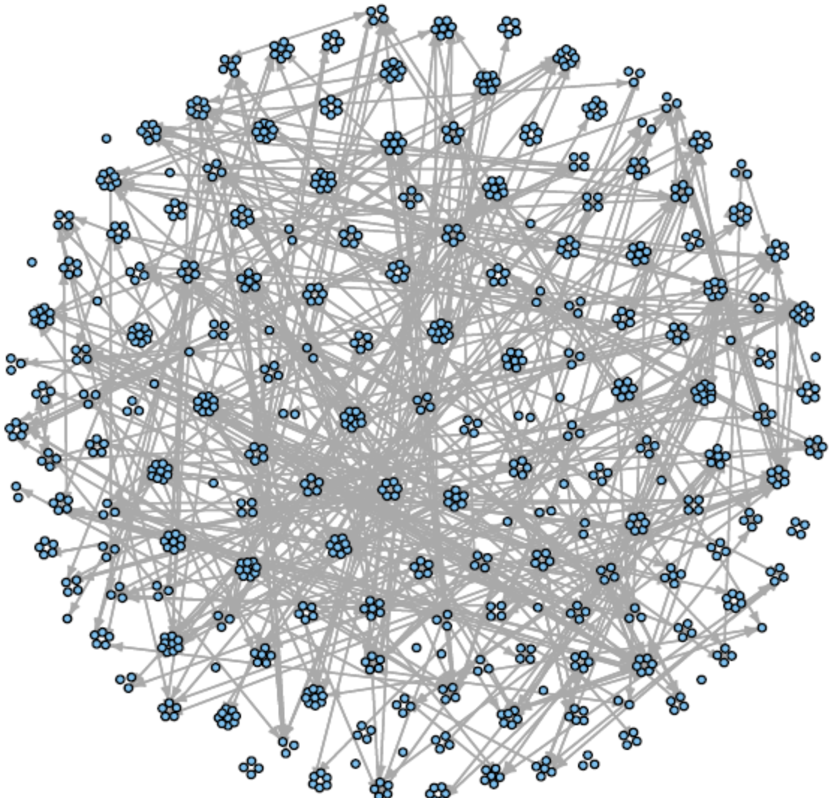

Advice



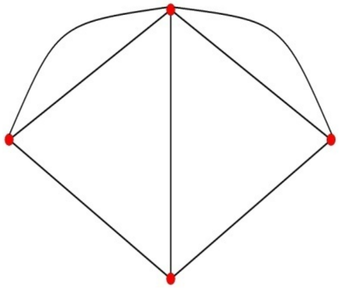
Kero−Come



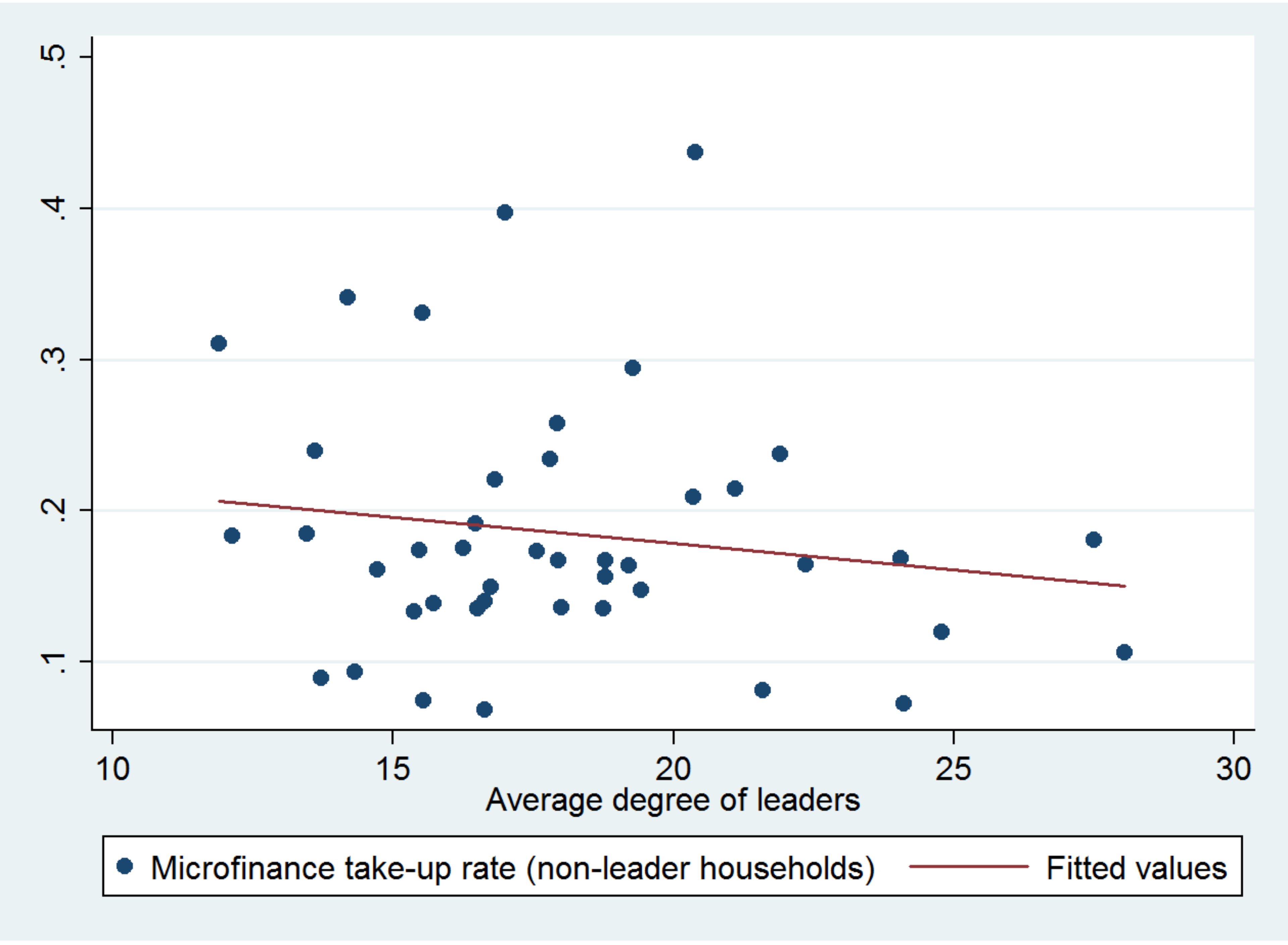
Medic



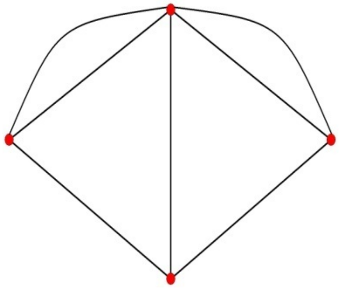
Centrality and diffusion



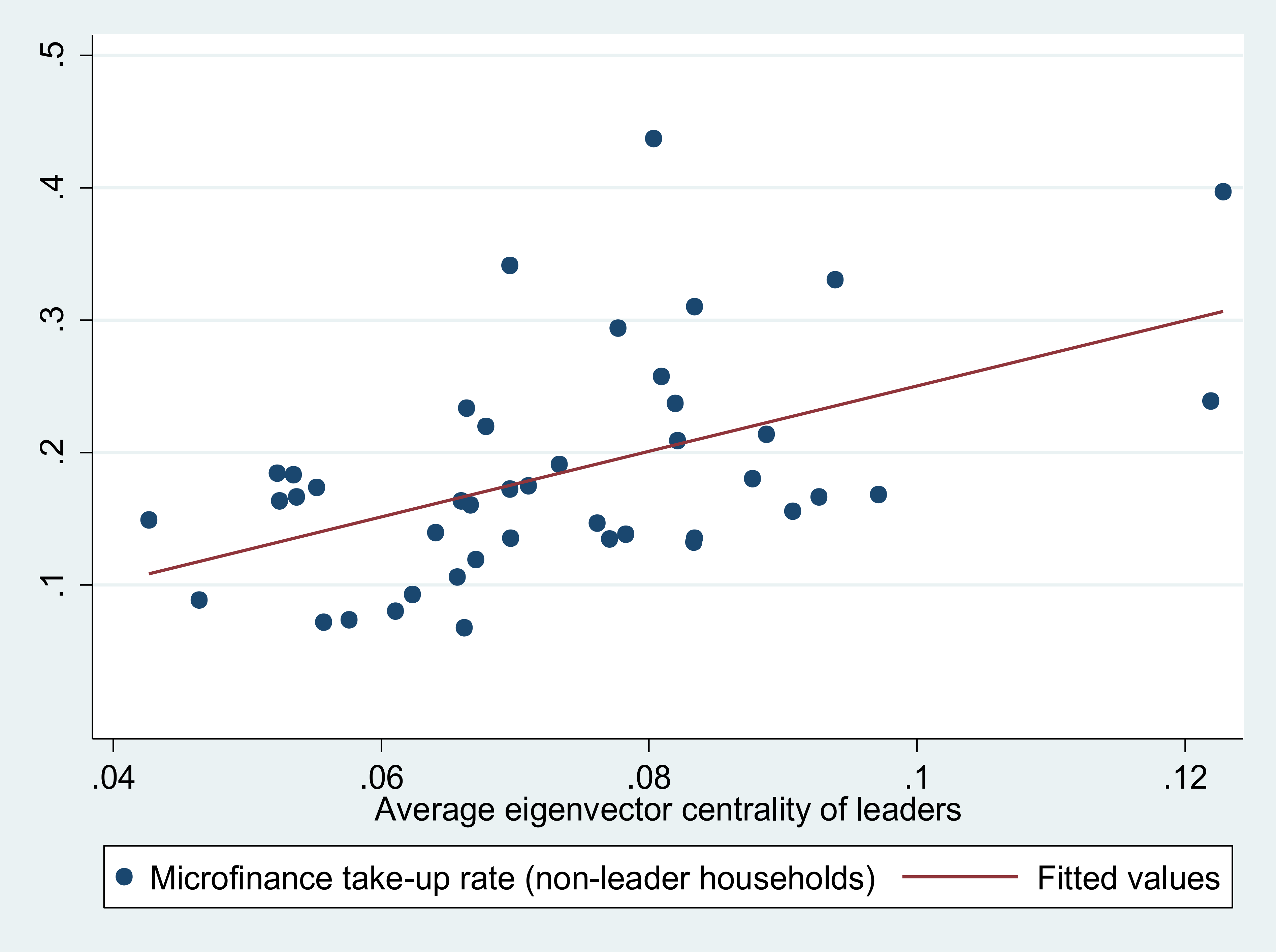
- Examine how centrality of ``injection points'' correlates with eventual diffusion of microfinance
- First up, degree centrality


.5
.4
.3
.2
.1
10
15
20
25
30
Average degree of leaders
Microfinance take-up rate (non-leader households)
Fitted values

Hypothesis Revised



- In villages where first contacted people have higher eigenvector centrality, there should be more diffusion


.5
.4
.3
.2
.1
.04
.06
.08
.1
.12
Average eigenvector centrality of leaders
Microfinance take-up rate (non-leader households)
Fitted values

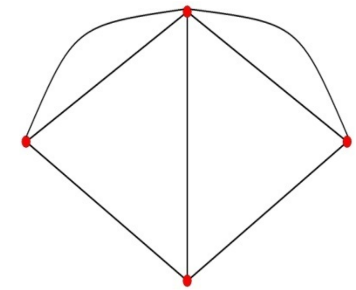
	Regress MF on				
Centrality:					
Eigen	1.723*				
	(.984)				
Degree		.177			
		(.118)			
Closeness			.804		
			(.481)		
Bonacich				.024	
				(.030)	
Between					.046
					(.032)
Obs	43	43	43	43	43
R-square	.324	.314	.309	.278	.301

Covariates: numHH, SHG, Savings, fracGM

	MF					
Centrality:						
DC	.429***					
	(.127)					
Eigen		1.723*				
		(.984)				
Degree			.177			
			(.118)			
Closeness				.804		
				(.481)		
Bonacich					.024	
					(.030)	
Between						.046
						(.032)
Obs	43	43	43	43	43	43
R-square	.470	.324	.314	.309	.278	.301

Covariates: numHH, SHG, Savings, fracGM

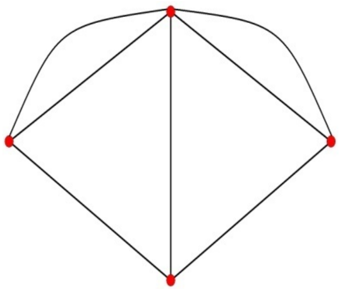
Summary so far:



- Networks are prevalent and important in many interactions (labor markets, crime, garment industry, risk sharing...)

- Although complex, social networks have identifiable characteristics:
  - ``small'' average and maximum path length
  - degree distributions that exhibit different shapes
  - homophily – strong tendency to associate with own type
  - centrality measures can capture node influence

Diffusion Centrality: DC_i (p,T)



- How many nodes end up informed if:

  - *i* is initially informed,
  - each informed node tells each of its neighbors with prob p in each period,
  - run for T periods?