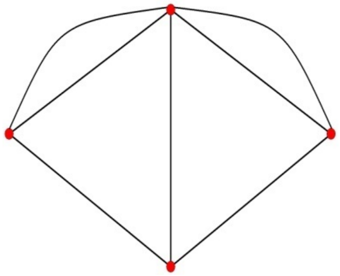
Social and Economic Networks: Models and Analysis



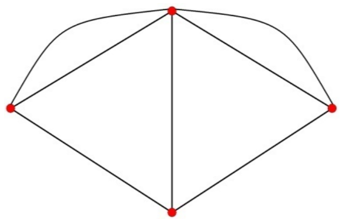
Matthew O. Jackson

Stanford University,

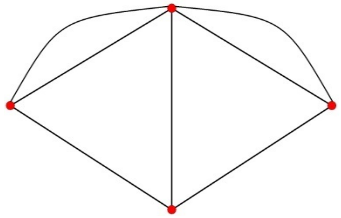
Santa Fe Institute, CIFAR,

www.stanford.edu\~jacksonm

2.1: Homophily

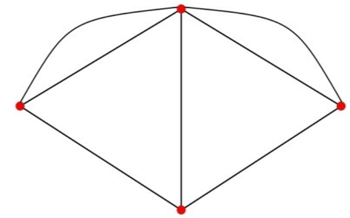


Outline



- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

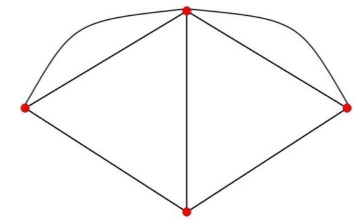
Homophily:



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

- age, race, gender, religion, profession….
  - Lazarsfeld and Merton (1954) ``Homophily''
  - Shrum (gender, ethnic, 1988…), Blau (professional 1964, 1977), Marsden (variety, 1987, 1988), Moody (grade, racial, 2001…), McPherson (variety,1991…)…

Illustrations Homophily:



- National Sample: only 8% of people have *any* people of another race that they ``discuss important matters'' with (Marsden 87)

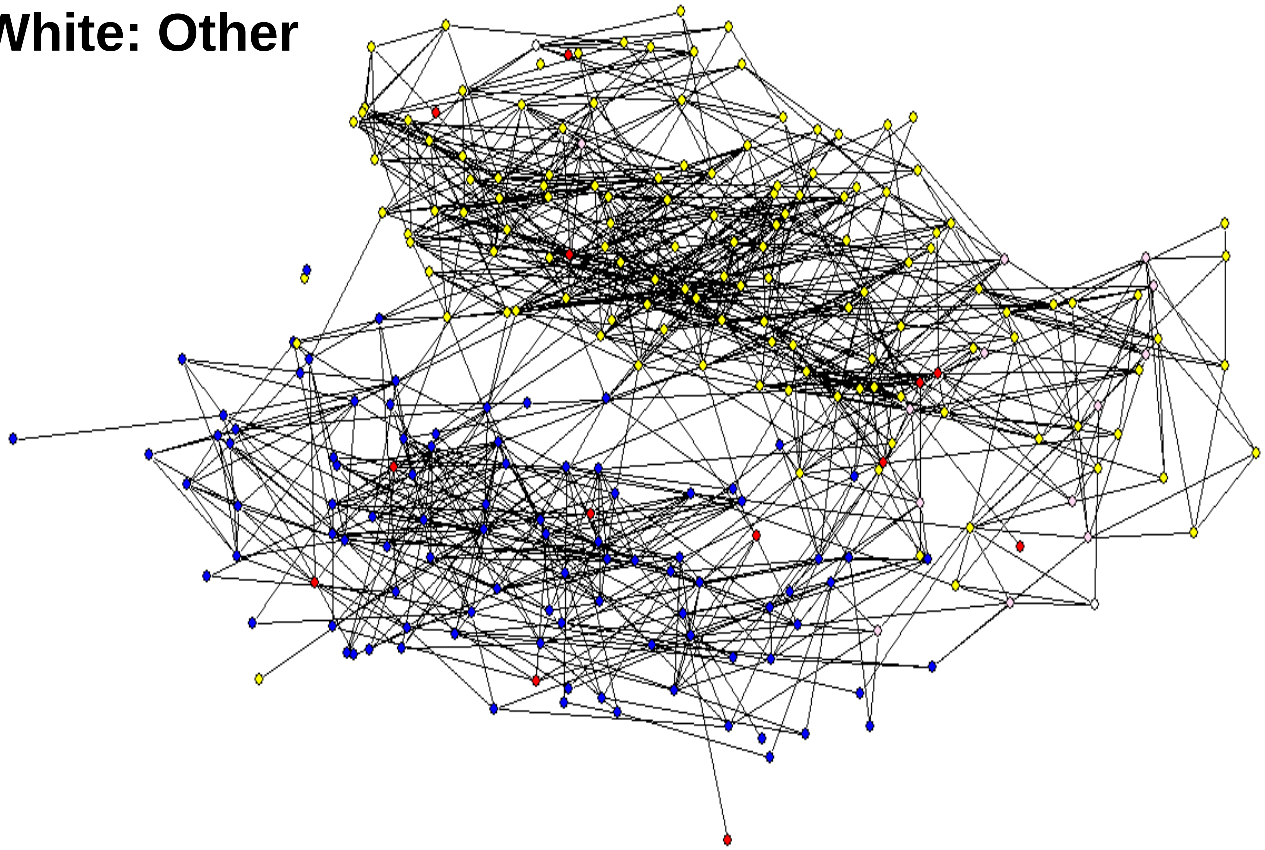
- Interracial marriages U.S.: 1% of white marriages, 5% of black marriages, 14% of Asian marriages (Fryer 07)

- In middle school, less than 10% of ``expected'' cross-race friendships exist (Shrum et al 88)

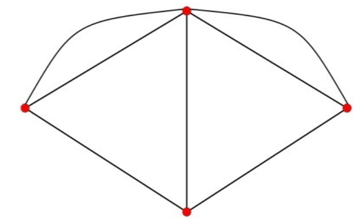
- Closest friend: 10% of men name a woman, 32% of women name a man (Verbrugge 77)

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

Currarini, Jackson, Pin 09,10



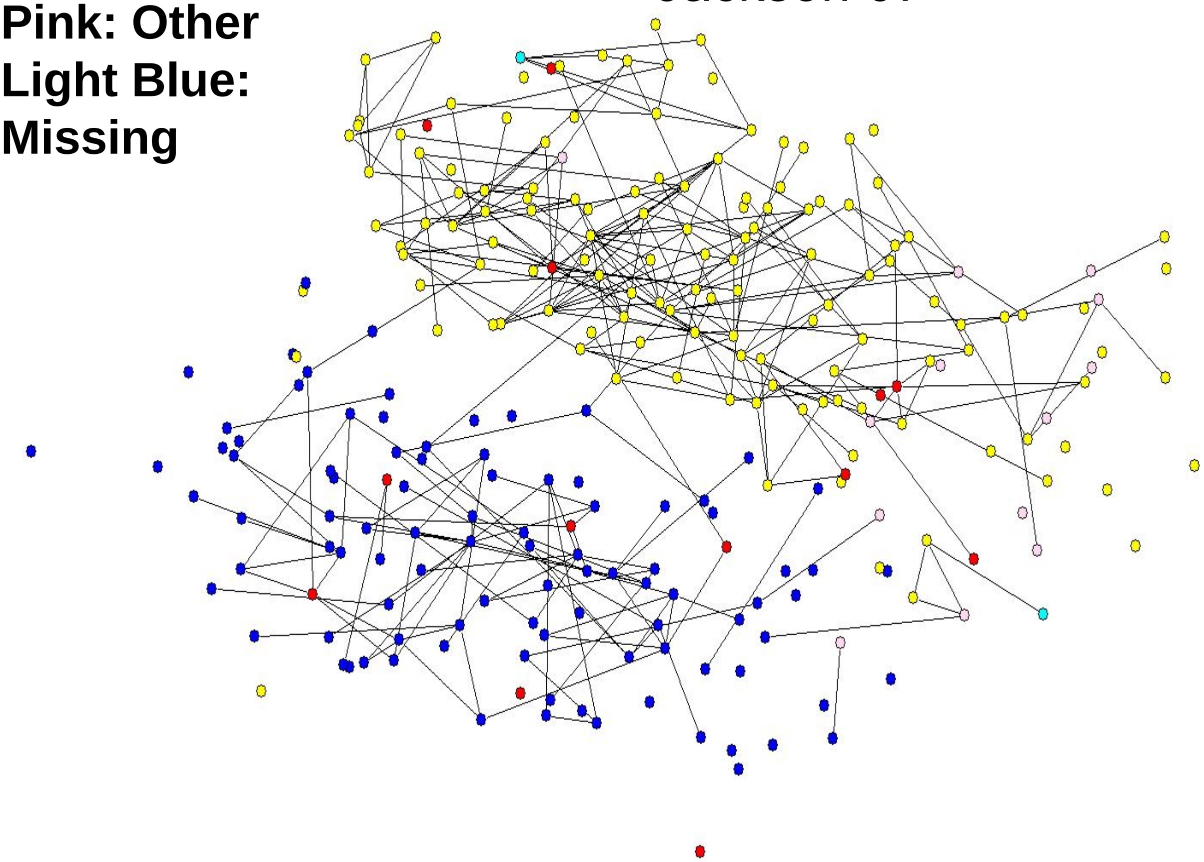
Adolescent Health, High School in US:



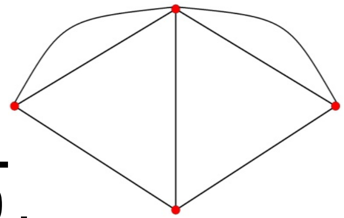
Percent:	52	38	5	5
	White	Black	Hispanic	Other
White	86	7	47	74
Black	4	85	46	13
Hispanic	4	6	2	4
Other	6	2	5	9
	100	100	100	100

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

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

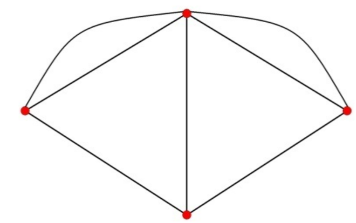


Baerveldt et al (2004) Homophily:



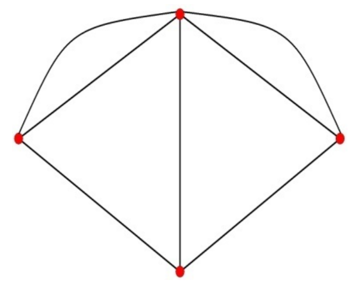
	n=850 65%	n=62 5%	n=75 6%	n=100 7%	n=230 17%
	Dutch	Moroccan	Turkish	Surinamese	Other
Dutch	79	24	11	21	47
Moroccan	2	27	8	4	5
Turkish	2	19	59	8	6
Surinamese	3	8	8	44	12
Other	13	22	14	23	30
	100	100	100	100	100

Reasons for Homophily



- opportunity – contact theory
- benefits/costs
- social pressure
- social competition...

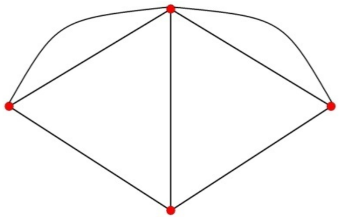
Social and Economic Networks: Models and Analysis



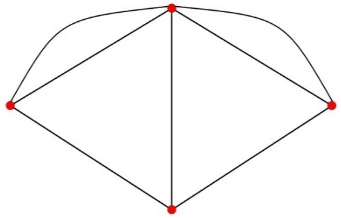
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.2: Dynamics and Tie Strength

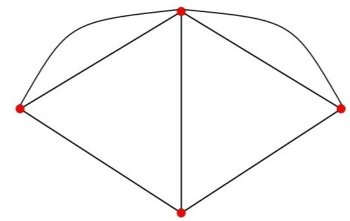


Outline



- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

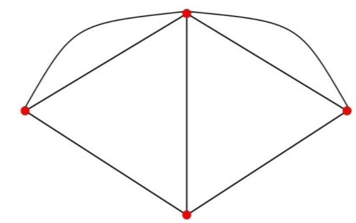
Dynamics



http://www.soc.duke.edu/~jmoody77/NetMovies/rom_flip.htm

http://www.soc.duke.edu/~jmoody77/NetMovies/soc_coath.htm

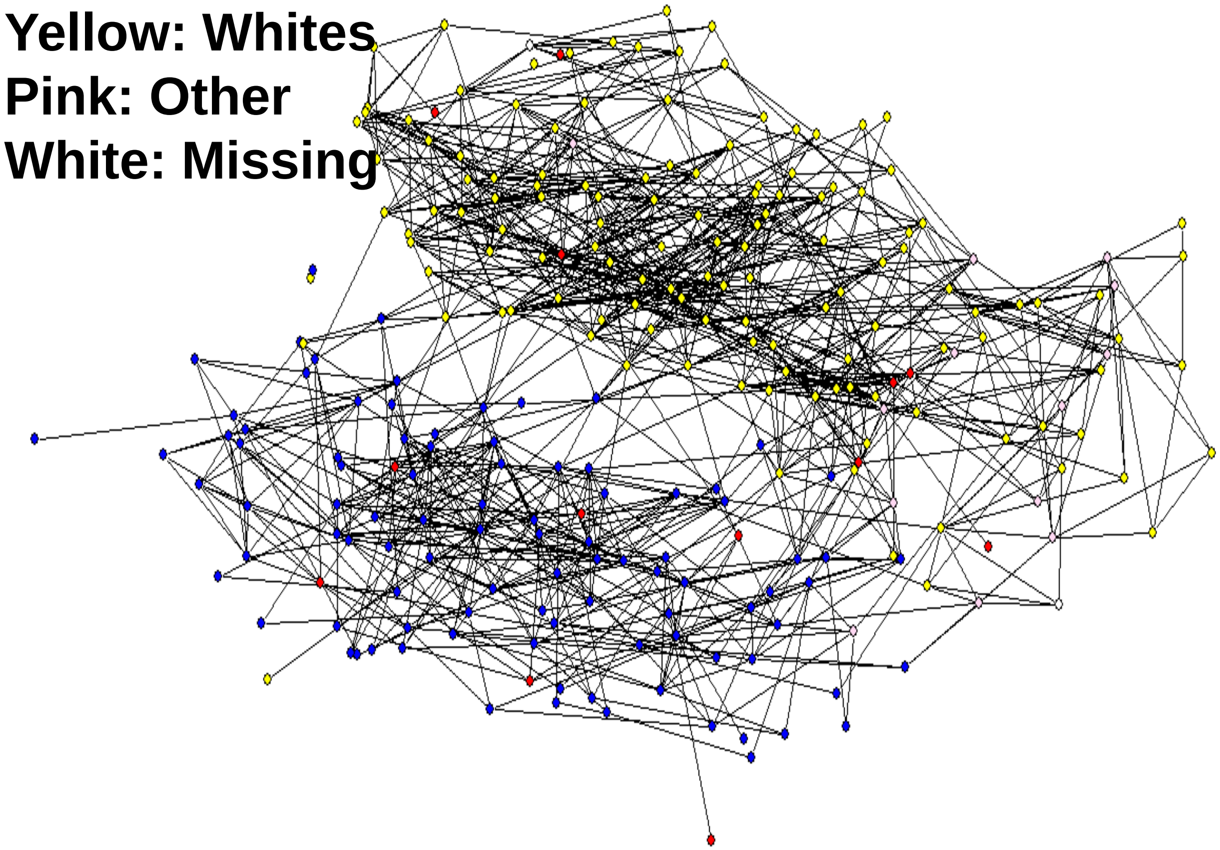
Strength of Weak Ties



- Granovetter interviews: 54 people who found their jobs via social tie:
  - 16.7 percent via strong tie (at least two interactions/week)
  - 55.7 percent via medium tie (at least one interaction per year)
  - 27.6 percent via a weak tie (less than one interaction per year)
- Theory: weak ties form `bridges', less redundant information

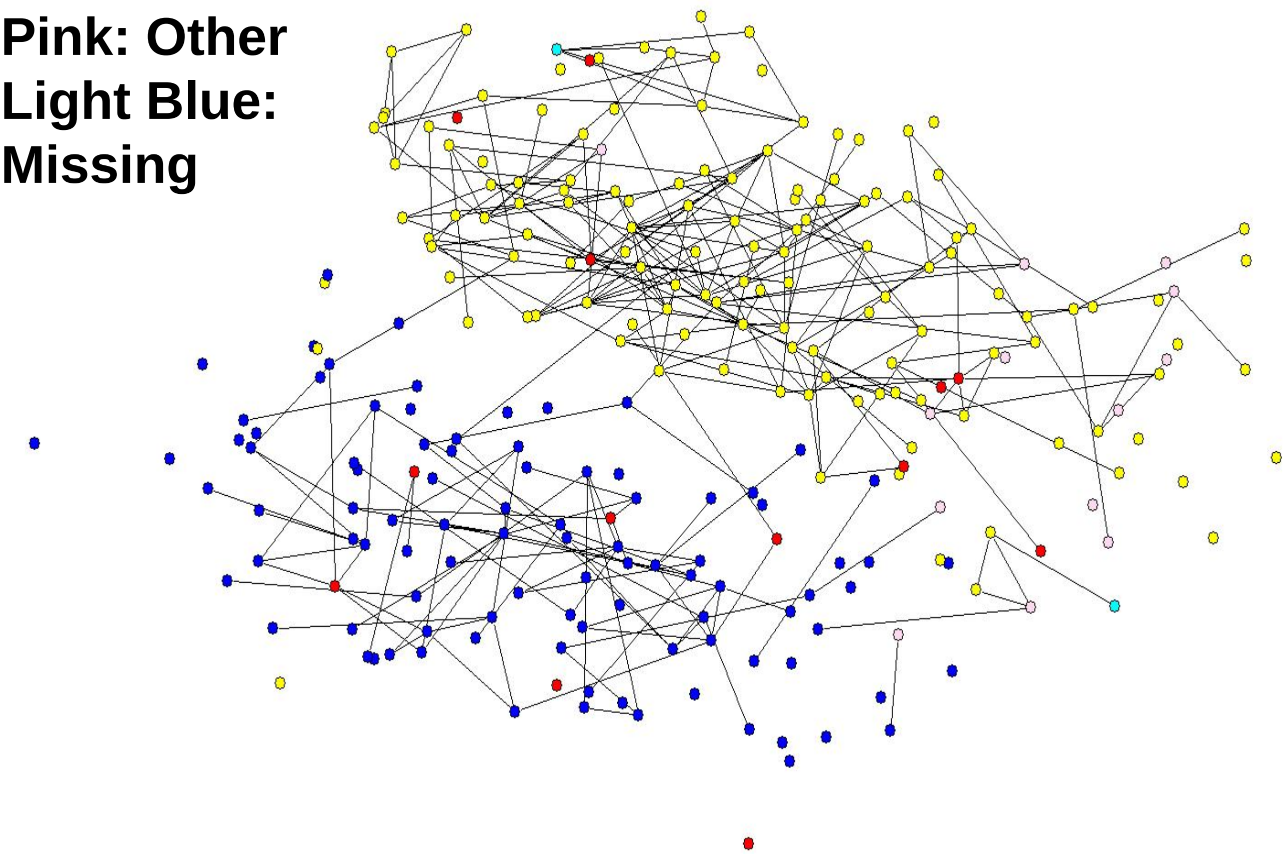
Add Health – from Currarini, Jackson, Pin (09,10)

Green: Asian
Blue: Blacks
Reds: Hispanics
Yellow: Whites
Pink: Other
White: Missing

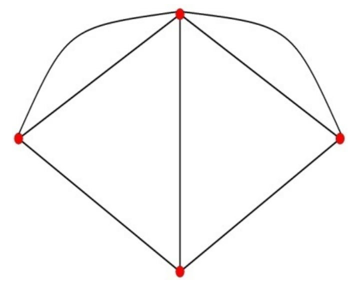


Blue: Black
Reds: Hispanic
Yellow: White
Pink: Other
Light Blue: Missing

"strong friendships"
cross group links less than half as frequent



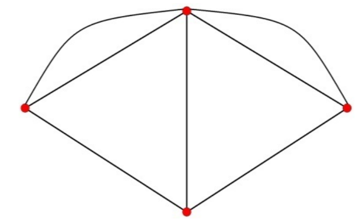
Social and Economic Networks: Models and Analysis



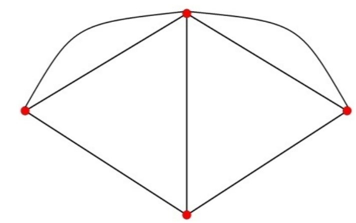
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.3: Centrality Measures – Degree, Closeness, Decay, and Betweenness

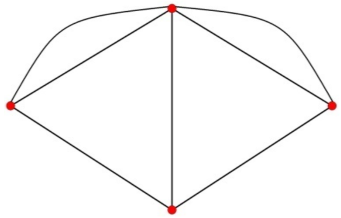


Simplifying the Complexity



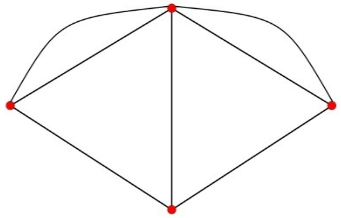
- Global patterns of networks
  - degree distributions, path lengths...
- Segregation Patterns
  - node types and homophily
- Local Patterns
  - Clustering, Transitivity, Support…
- Positions in networks
  - Neighborhoods, Centrality, Influence...

Position in Network



- How to describe individual characteristics?
  - Degree
  - Clustering
  - Distance to other nodes
  - Centrality, influence, power...???

Degree Centrality

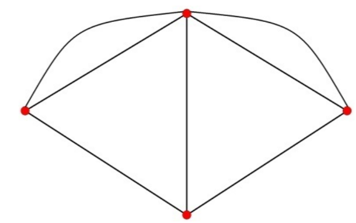


- How ``connected'' is a node?

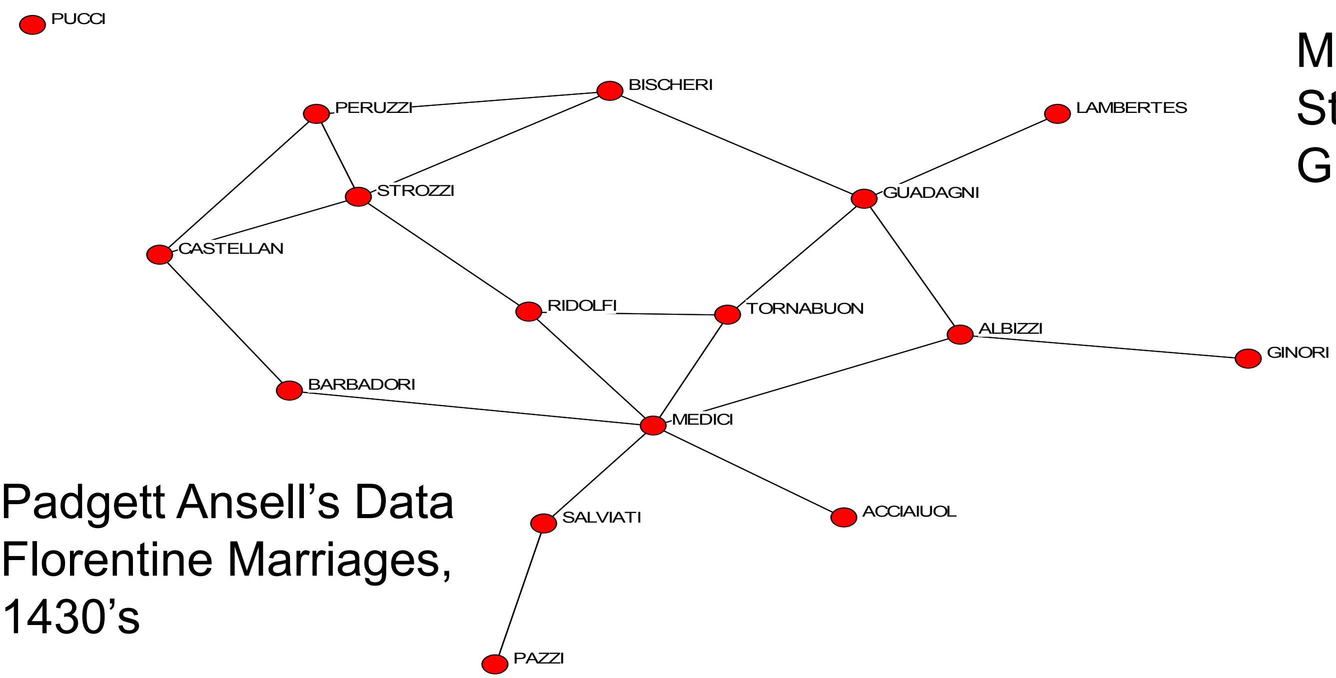
  - degree captures connectedness

  - normalize by n-1 - most possible

Degree Centrality

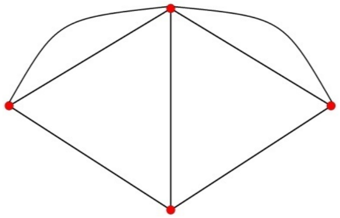


Medici = 6
Strozzi = 4
Guadagni = 4

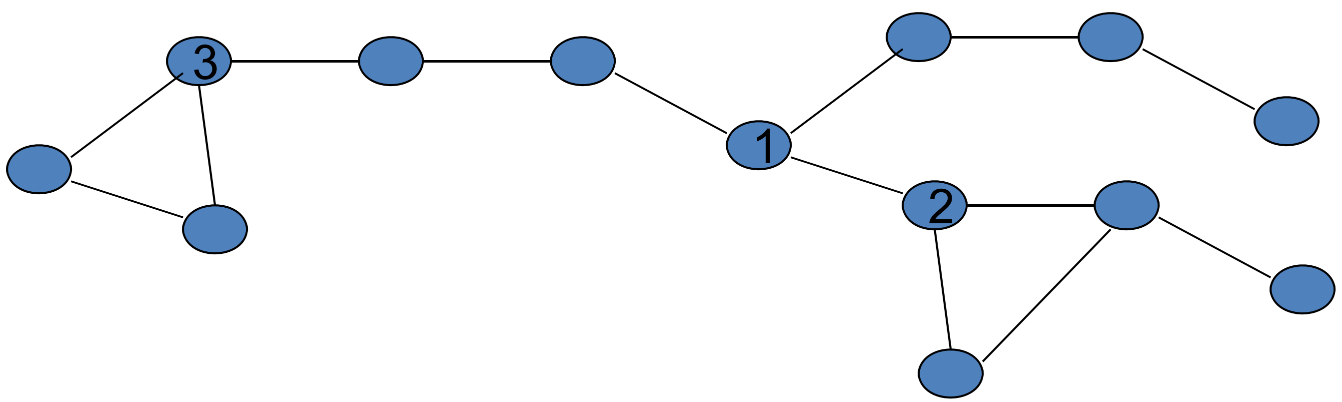


Padgett Ansell's Data
Florentine Marriages,
1430's

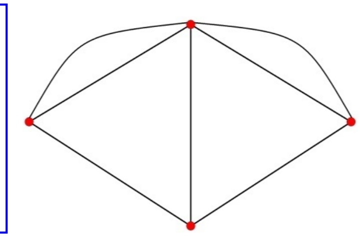
Degree Centrality



- Node 3 is considered as ``central'' as 1 and 2

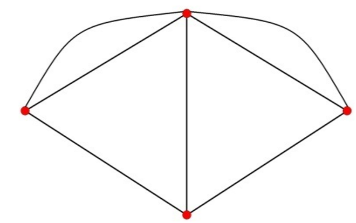


Centrality,
Four different things to measure:



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

Closeness

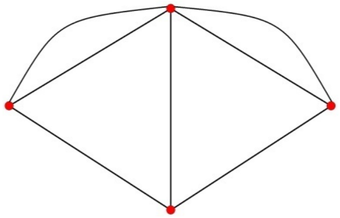


Closeness centrality: $(n-1) / \sum_j \boldsymbol{\ell}(i,j)$

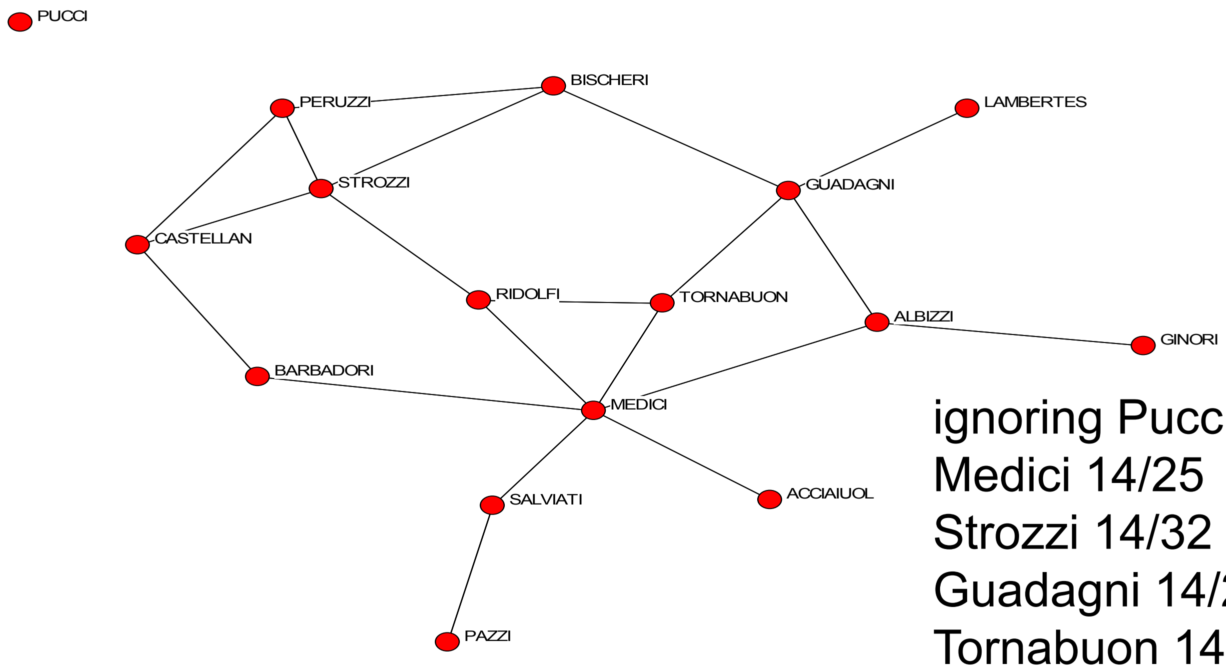
relative distances to other nodes

scales directly with distance – twice as far is half as central.

Closeness

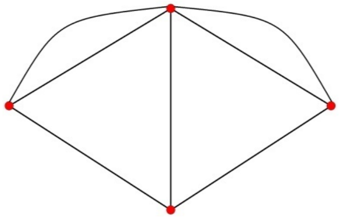


Closeness centrality: $(n-1) / \sum_j \ell(i,j)$



ignoring Pucci:
Medici 14/25
Strozzi 14/32
Guadagni 14/26
Tornabuon 14/29
Ridolfi 14/28

Decay Centrality



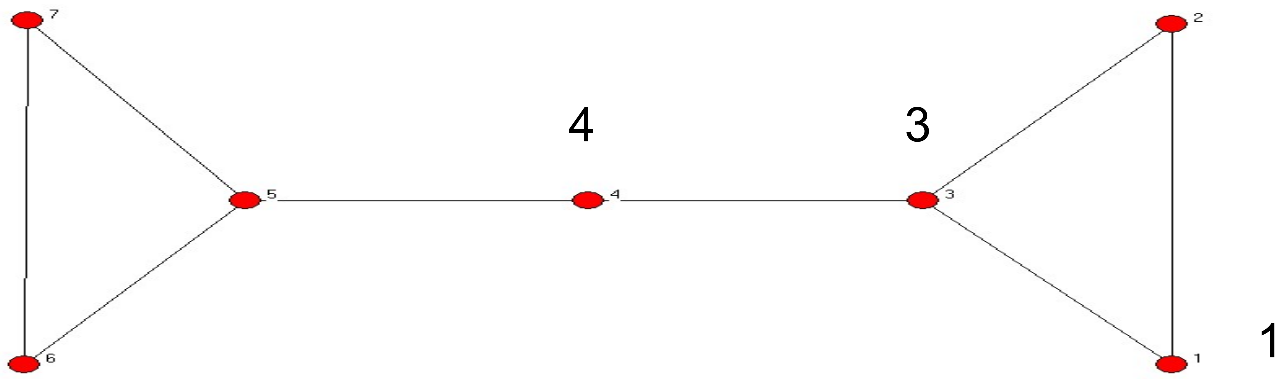
$C_i^d(g) = \sum_{j \neq i} \delta^{\ell(i,j)}$

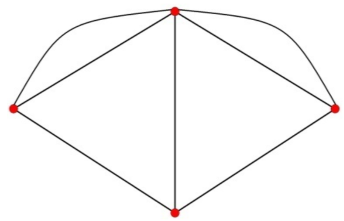
δ near 1 becomes component size

δ near 0 becomes degree

δ in between decaying distance measure

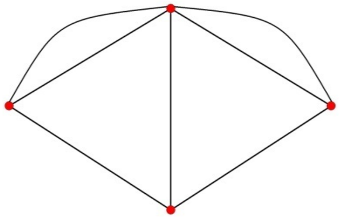
– weights distance exponentially





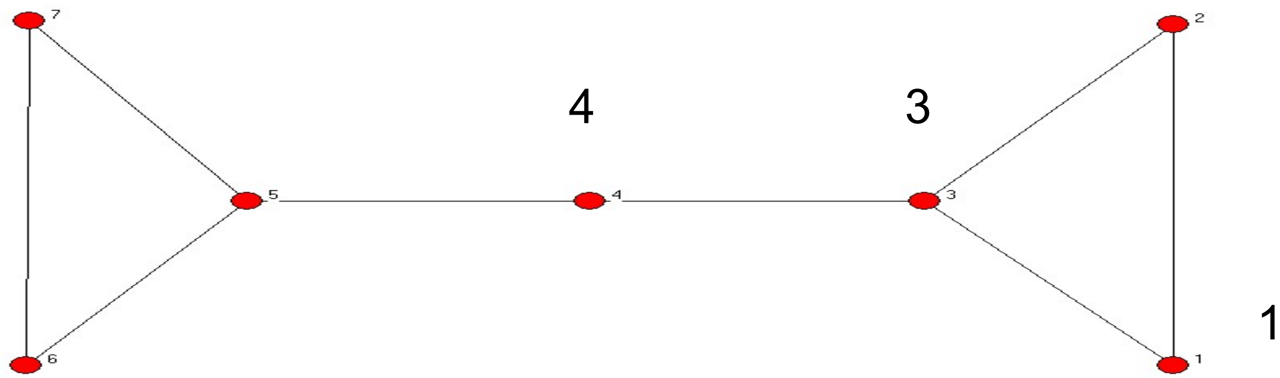
	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
Decay δ = .5	1.5	2.0	2.0
Decay δ = .75	3.1	3.7	3.8
Decay δ = .25	.59	.84	.75

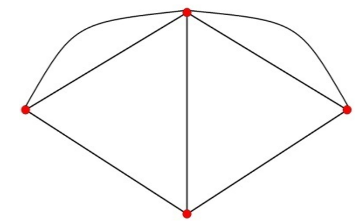
Normalize: Decay Centrality



$$C_i^d(g) = \sum_{j \neq i} \delta^{\ell(i,j)} \,/\, ((n-1)\,\delta)$$

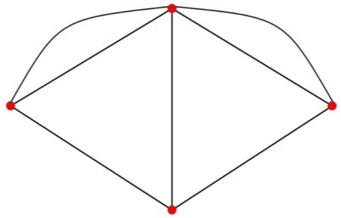
- $(n-1)\,\delta$ is the lowest decay possible





	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
N. Decay δ = .5	.50	.67	.67
N. Decay δ = .75	.69	.82	.84
N. Decay δ = .25	.39	.56	.50

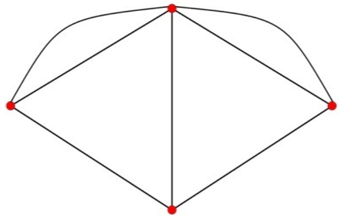
Betweenness (Freeman) Centrality

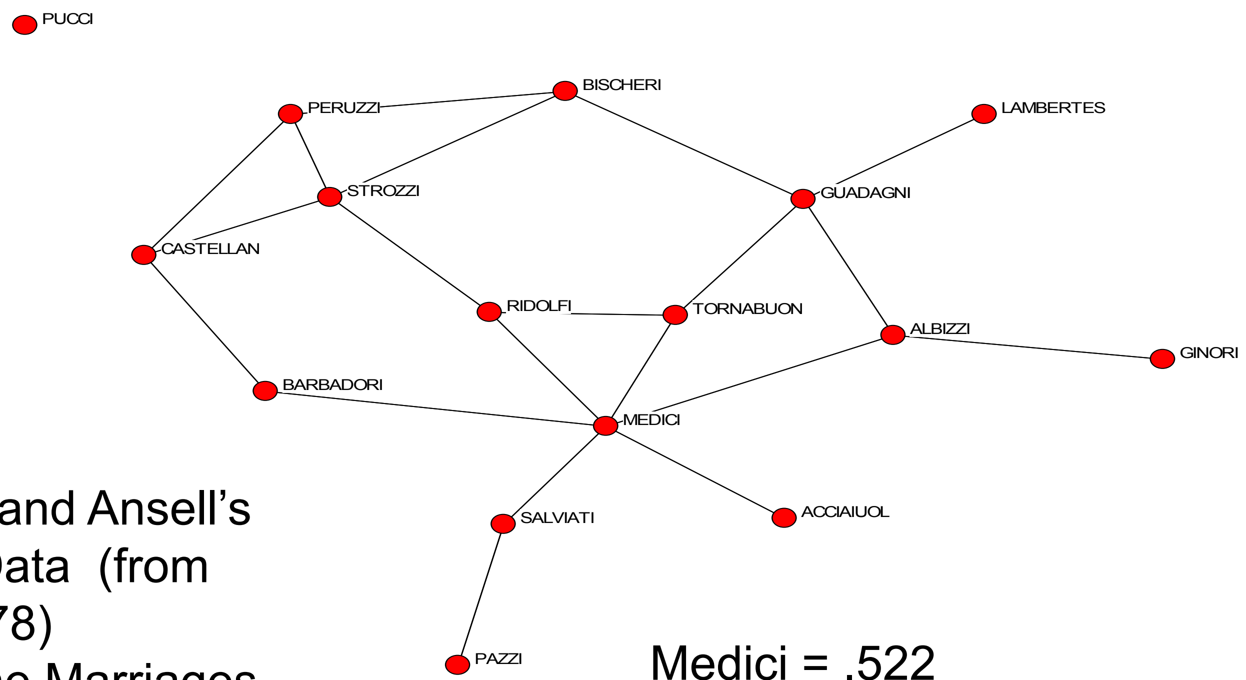


- $P(i,j)$ number of geodesics btwn i and j
- $P_k(i,j)$ number of geodesics btwn i and j that k lies on

- $\sum_{i,j \neq k} [P_k(i,j)/ P(i,j)] / [(n-1)(n-2)/2]$

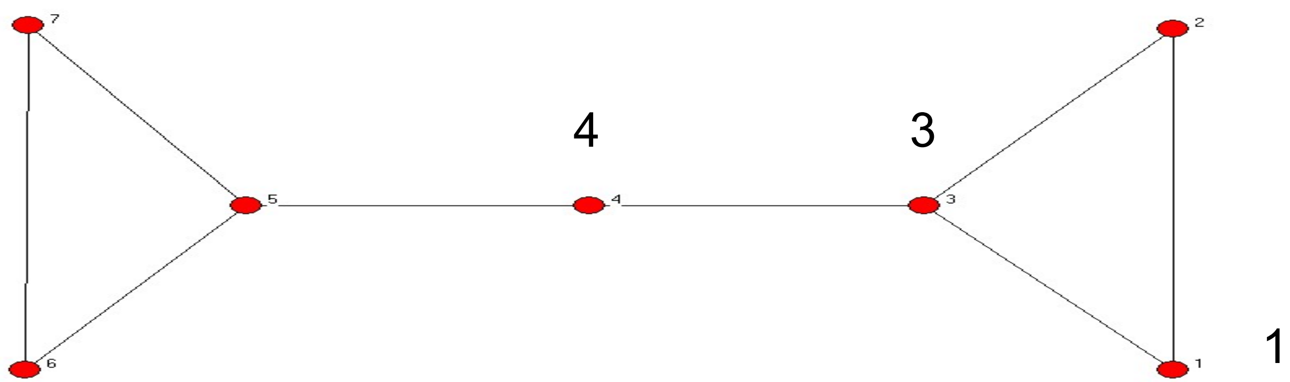
Betweenness Centrality

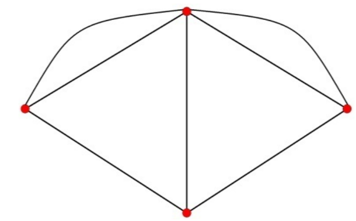




Padgett and Ansell's (1993) Data (from Kent 1978) Florentine Marriages, 1430's

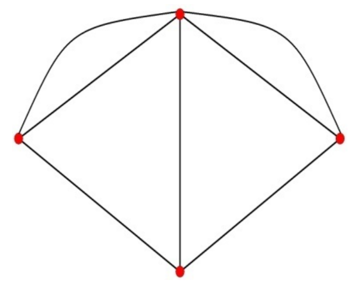
Medici = .522
Strozzi = .103
Guadagni = .255





	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
N. Decay δ = .5	.50	.67	.67
N. Decay δ = .75	.69	.82	.84
N. Decay δ = .25	.39	.56	.50
Betweenness	.00	.53	.60

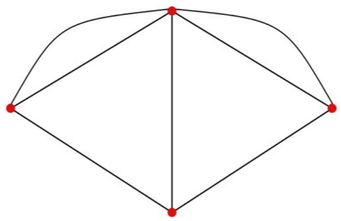
Social and Economic Networks: Models and Analysis



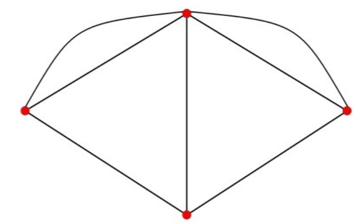
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.4: Centrality – Eigenvector Measures

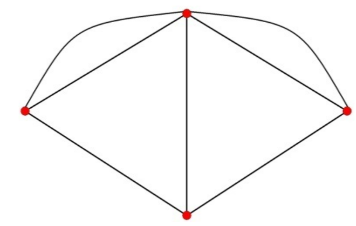


Position in Network

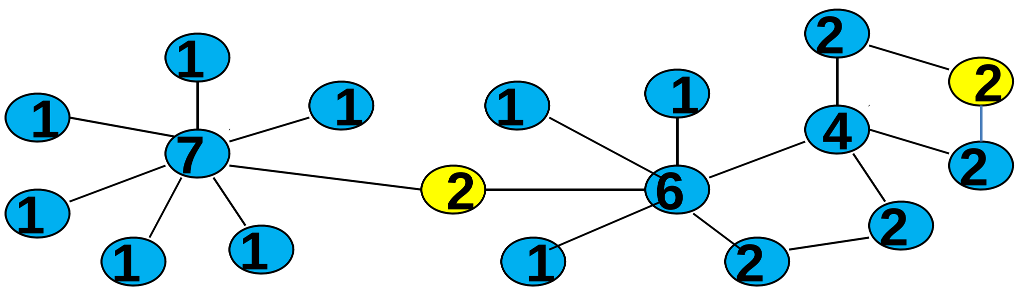


- How to describe individual characteristics?
  - Degree
  - Clustering
  - Distance to other nodes
  - Centrality, influence, power...???

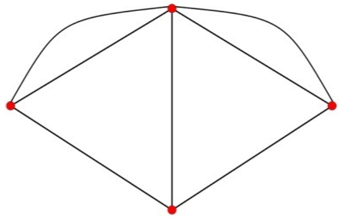
Degree Centrality?



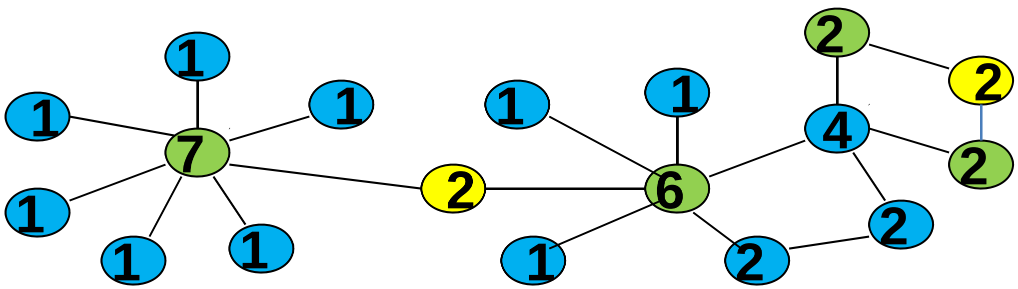
- Failure of degree centrality to capture reach of a node:



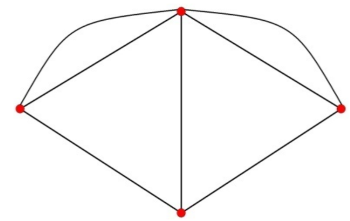
Degree Centrality?



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



Eigenvector Centrality

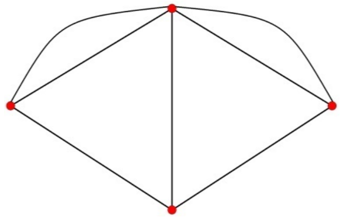


- Centrality is proportional to the sum of neighbors' centralities

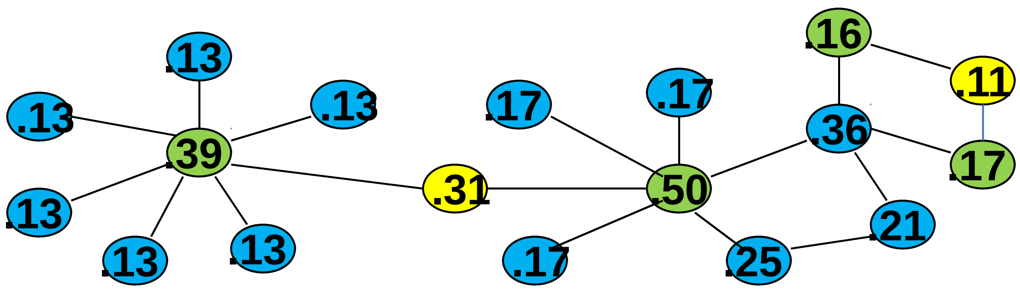
C_i proportional to $\sum_{j:\text{ friend of } i} C_j$

$C_i = a \sum_j g_{ij} C_j$

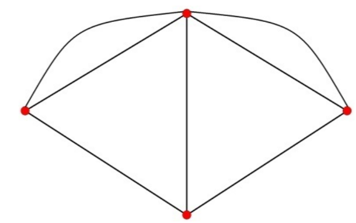
Eigenvector Centrality



Now distinguishes more ``influential'' nodes



Prestige, Influence, Eigenvector-based Centrality



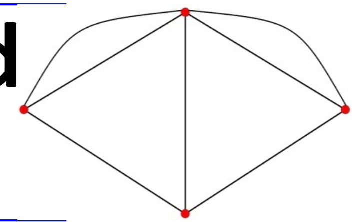
- Get value from connections to others, but proportional to their value

- Self-referential concept

$$C_i^e(g) = a \sum_j g_{ij} C_j^e(g)$$

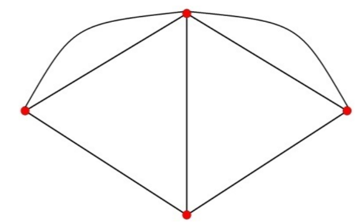
– centrality is proportional to the summed centralities of neighbors

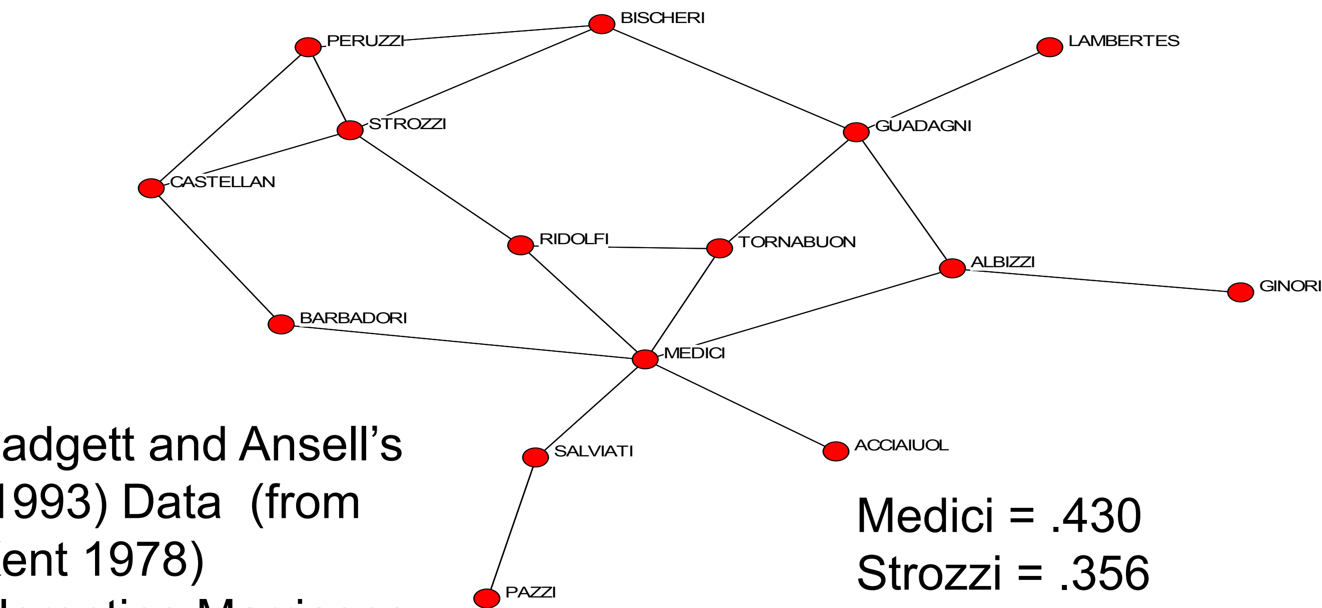
Prestige, Influence, Eigenvector-based Centrality



- $C_i^e(g) = a \sum_j g_{ij} C_j^e(g)$ $\qquad C^e(g) = a\, g\, C^e(g)$
  - $C^e(g)$ is an eigenvector - many possible solutions
  - Look for one with largest eigenvalue – will be nonnegative (Perron-Frobenius Theorem)
  - normalize entries to sum to one

Eigenvector Centrality

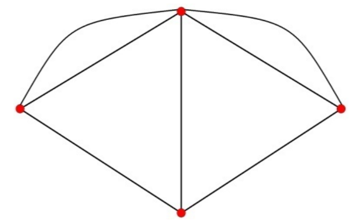




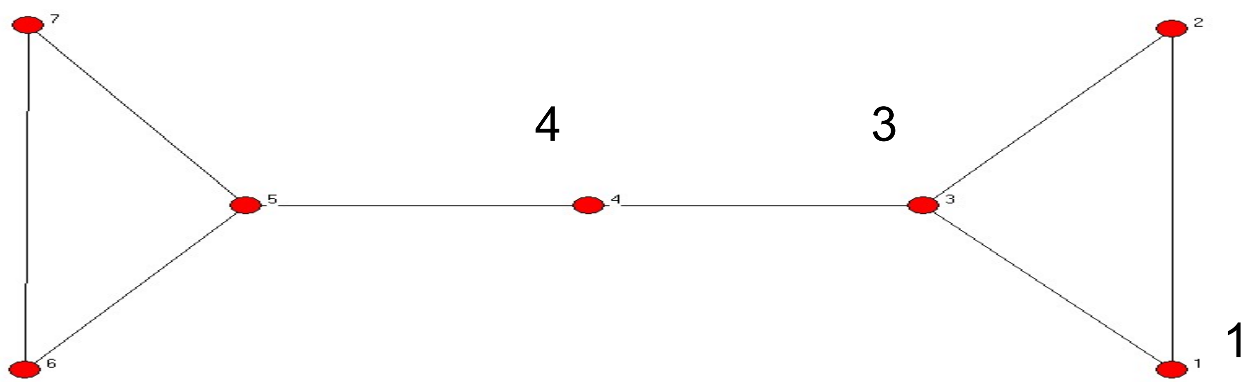
Padgett and Ansell's (1993) Data (from Kent 1978) Florentine Marriages, 1430's

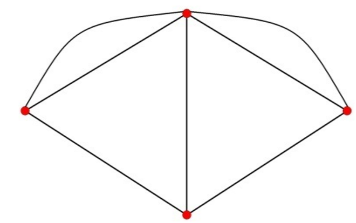
Medici = .430
Strozzi = .356
Guadagni = .289
Ridolfi=.341
Tornabuon=.326

Centrality



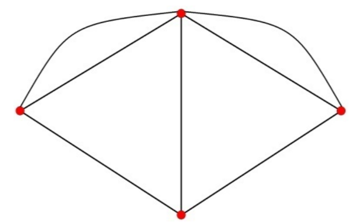
- Concepts related to eigenvector centrality:
- Google Page rank: score of a page is proportional to the sum of the scores of pages linked to it
- Random surfer model: start at some page on the web, randomly pick a link, follow it, repeat...





	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
N. Decay δ = .5	.50	.67	.67
N. Decay δ = .75	.69	.82	.84
N. Decay δ= .25	.39	.56	.50
Betweenness	.00	.53	.60
Eigenvector	.47	.63	.54

Bonacich Centrality



Builds on a measure by Katz

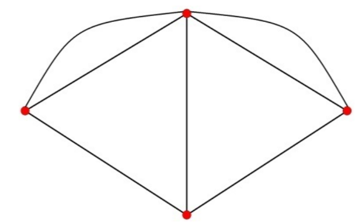
give each node a base value $ad_i(g)$ for some $a>0$

then add in all paths of length 1 from i to some j times b times j's base value

then add in all walks of length 2 from i to some j times b^2 times j's base value...

$$C^b(g) = ag\mathbf{1} + b\, g\, ag\mathbf{1} + b^2 g^2 ag\mathbf{1} \ldots$$

Bonacich Centrality



$$C^b(g) = ag\mathbf{1} + b\, g\, ag\mathbf{1} + b^2 g^2 ag\mathbf{1} \ldots$$

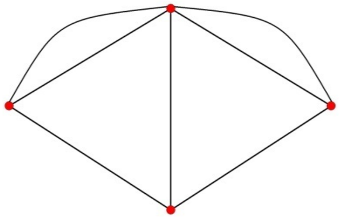
$$= a(g\mathbf{1} + b\, g^2\mathbf{1} + b^2 g^3\mathbf{1} \ \ldots)$$

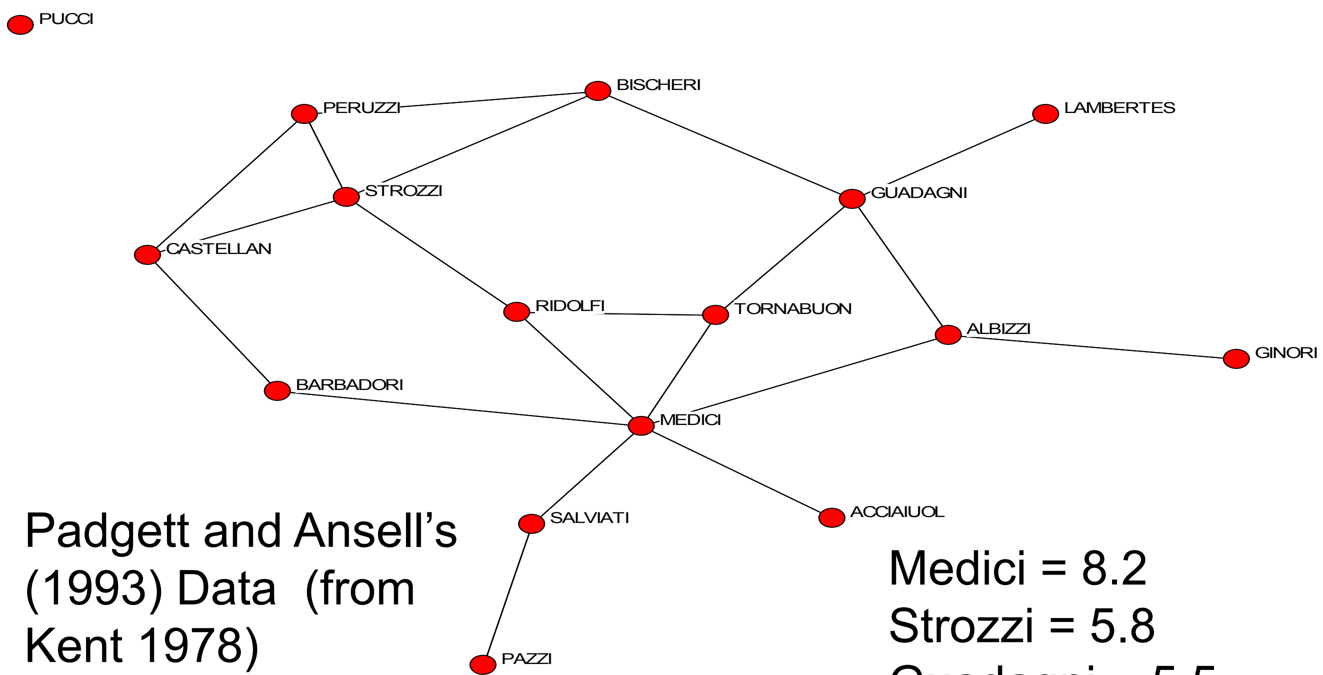
normalize a to 1, need *small b* to be finite

$$C^b(g) = g\mathbf{1} + b\, g^2\mathbf{1} + b^2 g^3\mathbf{1} \ \ldots$$

$$= (\mathbf{I} - bg)^{-1} g\mathbf{1}$$

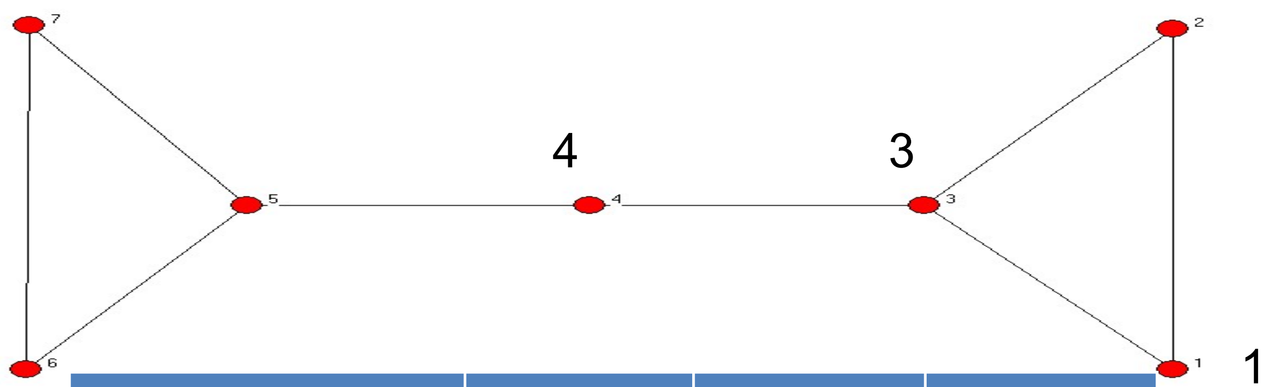
Bonacich Centrality

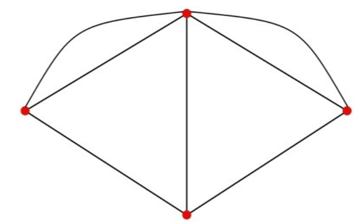




Padgett and Ansell's (1993) Data (from Kent 1978) Florentine Marriages, 1430's

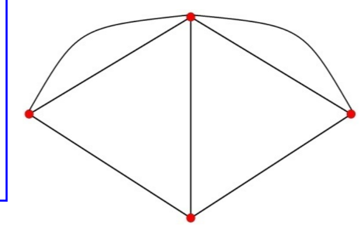
Medici = 8.2
Strozzi = 5.8
Guadagni = 5.5
Ridolfi=4.9
Tornabuon=4.9





	Node 1	Node 3	Node 4
Degree	.33	.50	.33
Closeness	.40	.55	.60
N. Decay δ= .5	.50	.67	.67
N. Decay δ = .75	.69	.82	.84
N. Decay δ = .25	.39	.56	.50
Betweenness	.00	.53	.60
Eigenvector	.47	.63	.54
Bonacich b=1/3	9.4	13	11
Bonacich b=1/4	4.9	6.8	5.4

Centrality,
Four different things to measure:



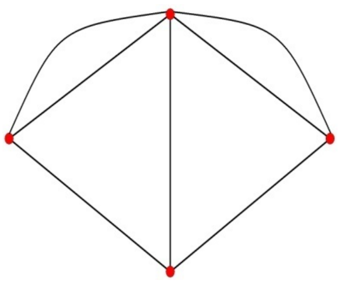
- Degree – connectedness

- Closeness, Decay – ease of reaching other nodes

- Betweenness – importance as an intermediary, connector

- Influence, Prestige, Eigenvectors – ``not what you know, but who you know..''

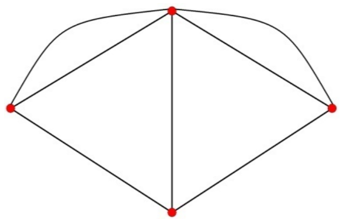
Social and Economic Networks: Models and Analysis



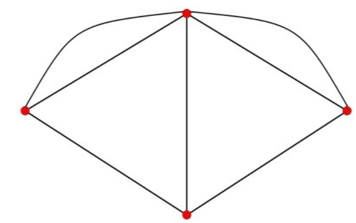
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.5a: Application – Centrality Measures

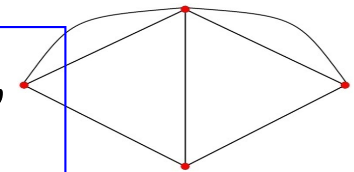


Centrality Application: What affects Diffusion?

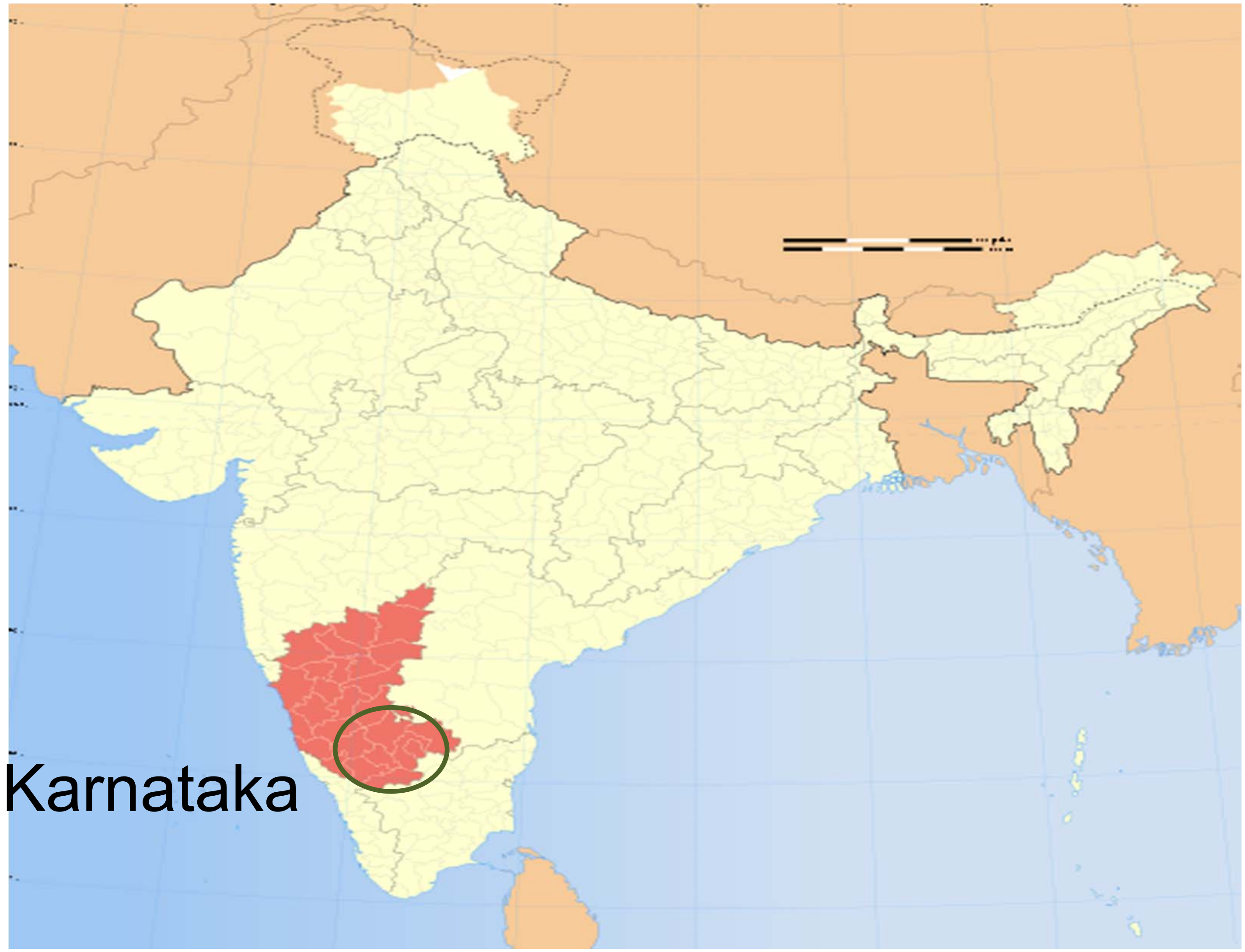


- First contact points: let us examine how network positions of injection points matter

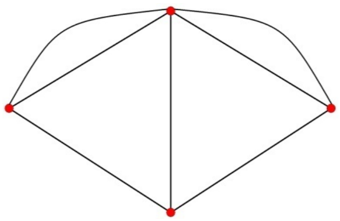
Banerjee, Chandrasekhar, Duflo, Jackson, Diffusion of Microfinance (2013)



- 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

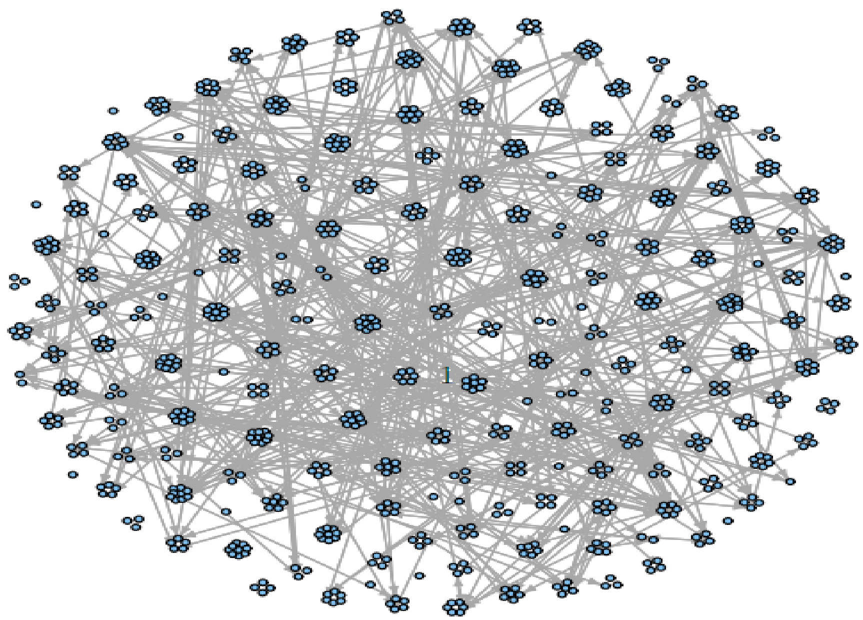

Karnataka

Background: 75 Indian Villages – Networks

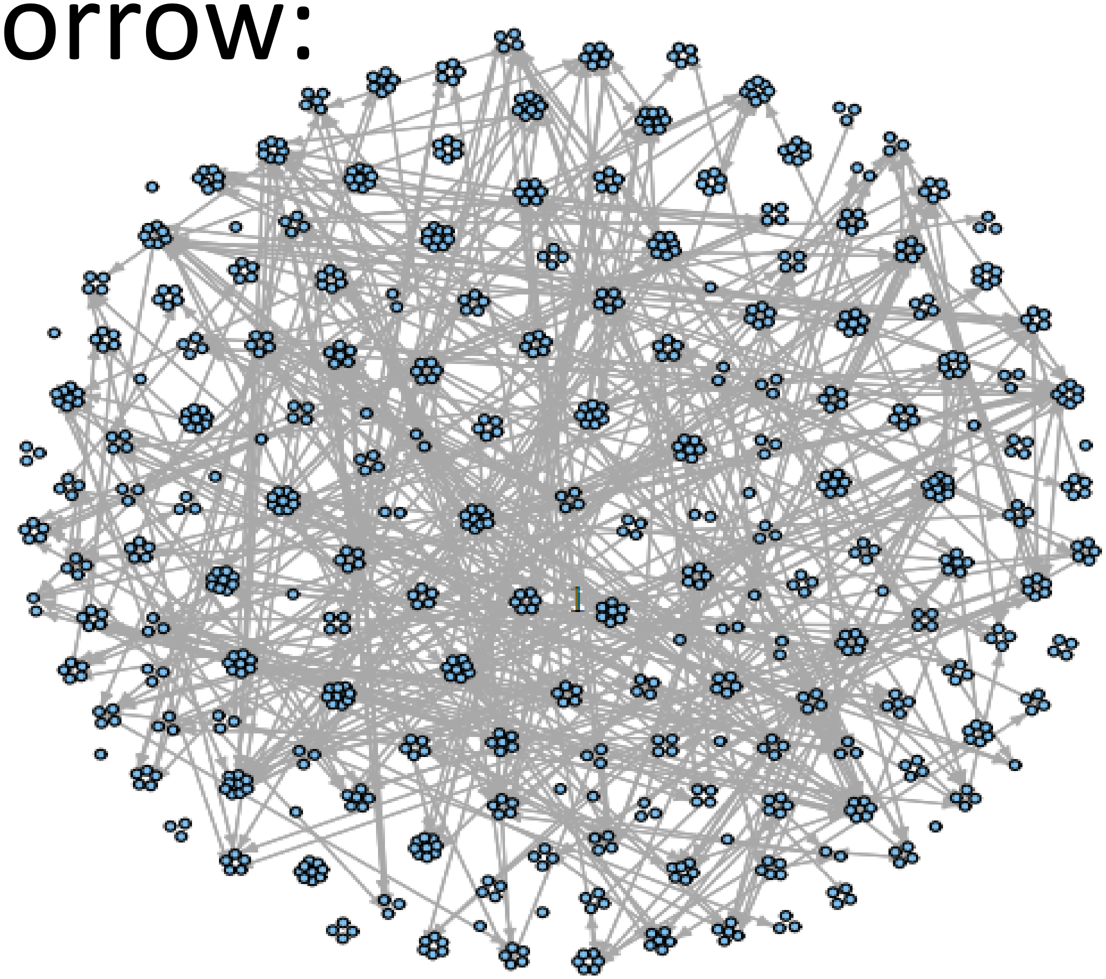


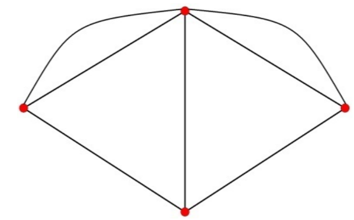
- ``Favor'' Networks:
  - both borrow and lend money
  - both borrow and lend kero-rice
- ``Social'' Networks:
  - both visit come and go
  - friends (talk together most)
- Others (temple, medical help…)

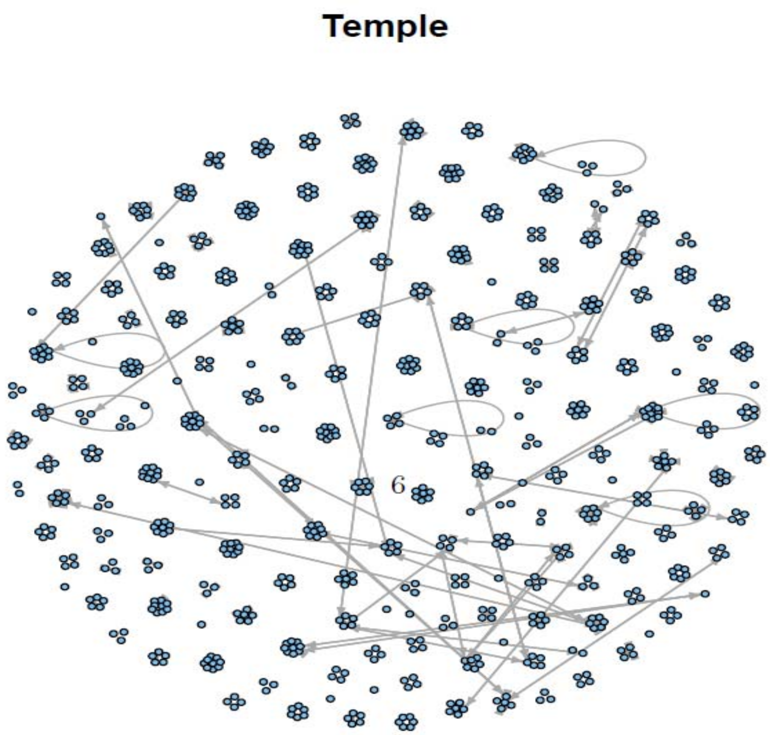
Borrow:

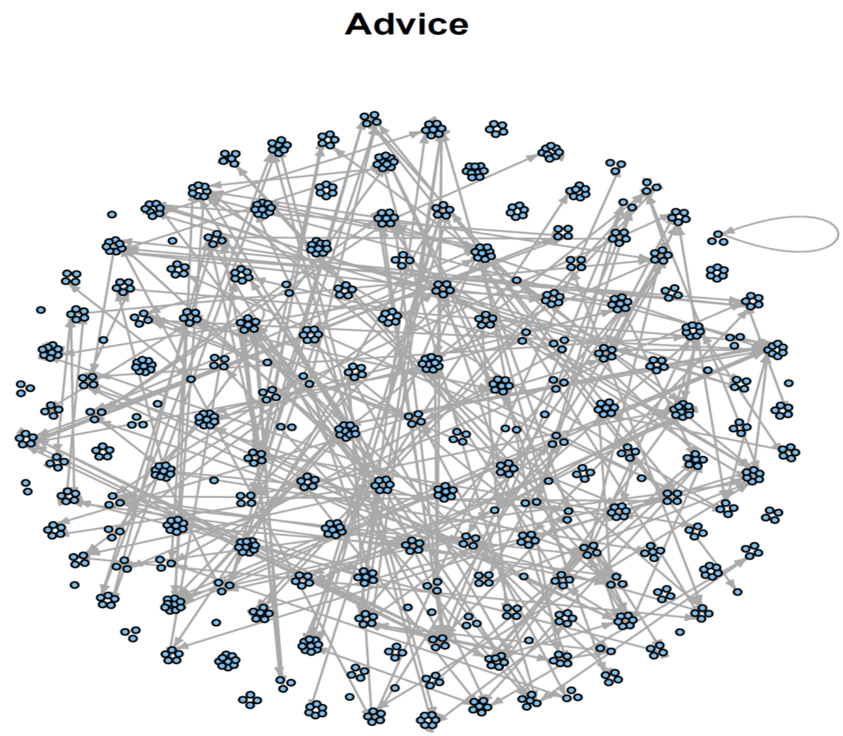


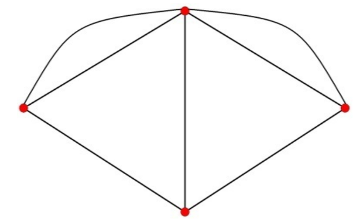
Borrow:



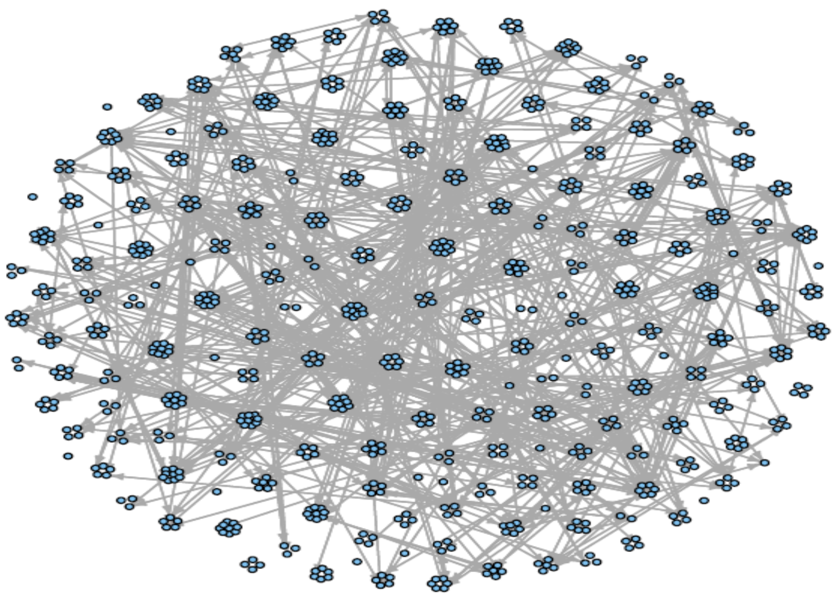



Temple
6

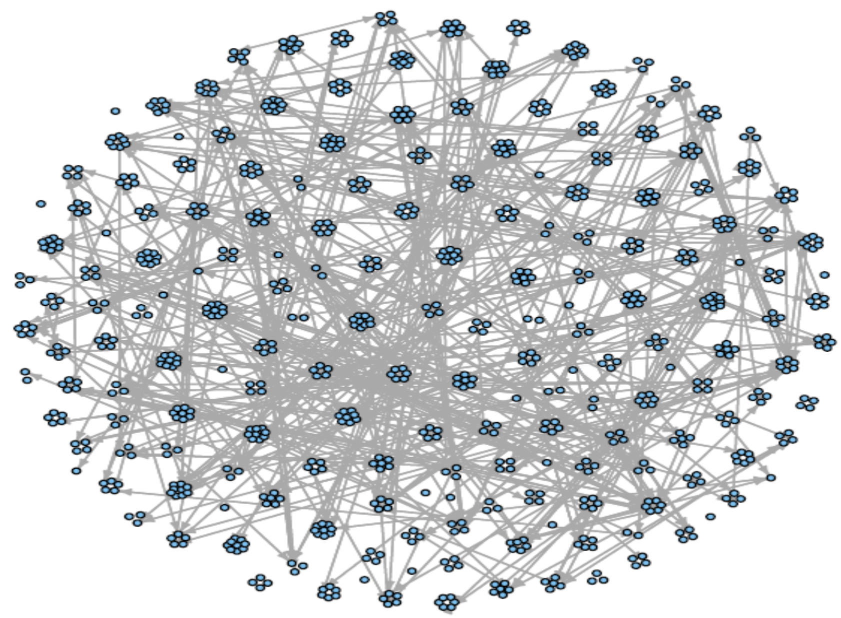

Advice



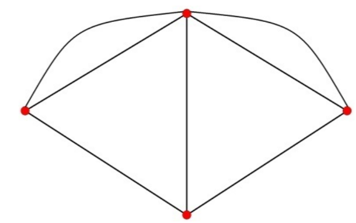
Kero−Come



Medic

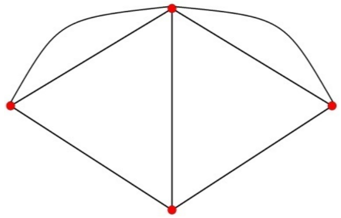


Data also include

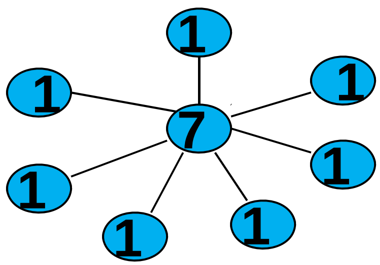


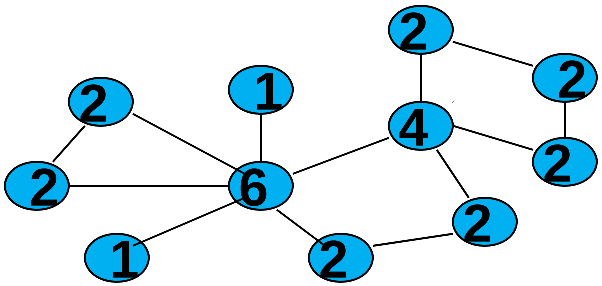
- Microfinance participation by individual, time
- Number of households and their composition
- Demographics: age, gender, subcaste, religion, profession, education level, family...
- Wealth variables: latrine, number rooms, roof,
- Self Help Group participation rate, ration card, voting
- Caste: village fraction of ``higher castes'' (GM/FC and OBC, remainder are SC/ST)

Degree Centrality

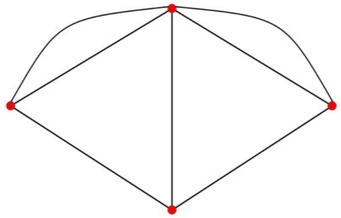


- Count how many links a node has



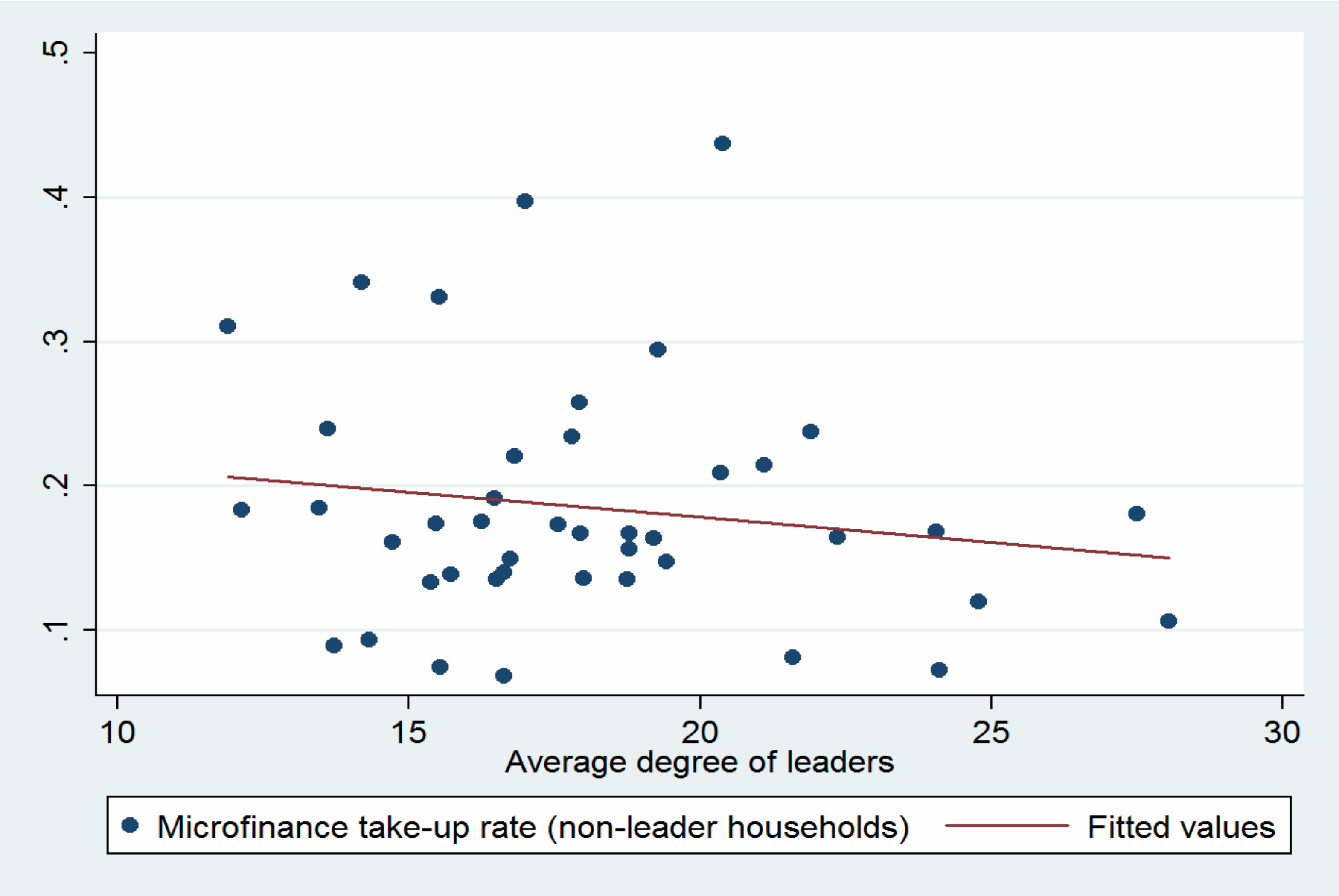


Hypothesis

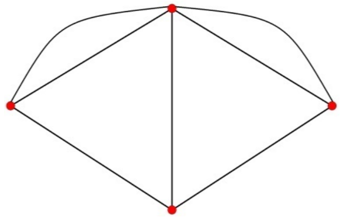


- In villages where first contacted people have more connections, there should be a better spread of information about microfinance

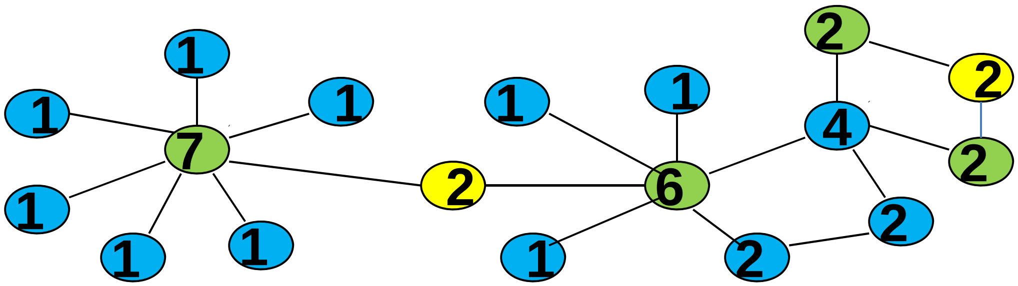
- more people knowing should lead to higher participation


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

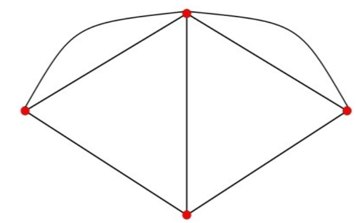
Degree Centrality?



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



Eigenvector Centrality

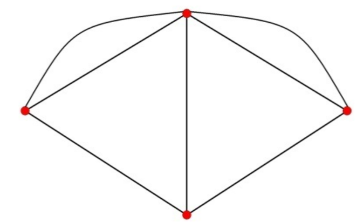


- Centrality is proportional to the sum of neighbors' centralities

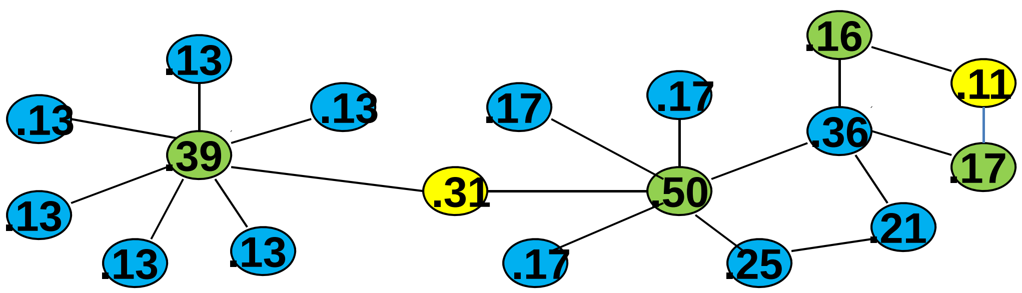
C_i proportional to $\sum_{j\text{: friend of } i} C_j$

$$C_i = a \sum_j g_{ij} C_j$$

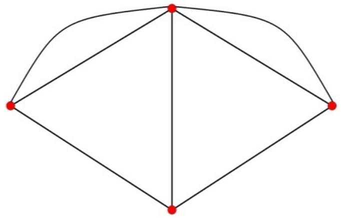
Eigenvector Centrality



Now distinguishes more ``influential'' nodes

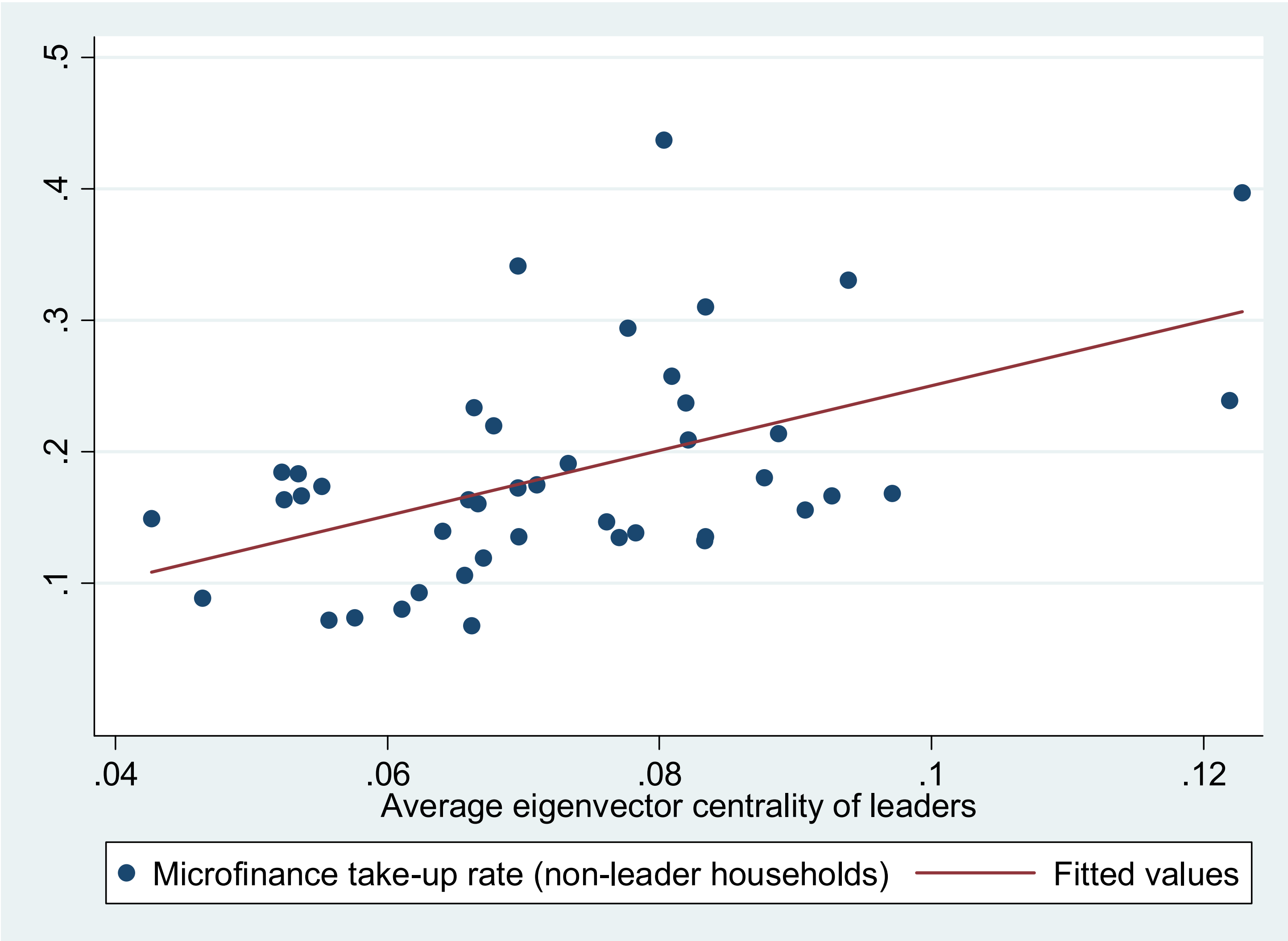


Hypothesis Revised



- In villages where first contacted people have higher eigenvector centrality, there should be a better spread of information about microfinance

- more people knowing should lead to higher participation

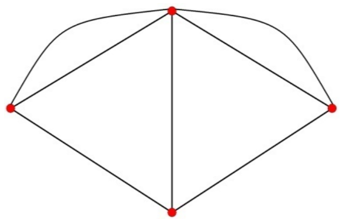

.1
.2
.3
.4
.5
.04
.06
.08
.1
.12
Average eigenvector centrality of leaders
Microfinance take-up rate (non-leader households)
Fitted values

VARIABLES	MF Participation
eigLeader	1.93**
	(0.93)
degreeLeader	-0.003
	(0.003)
numHH	-0.0003
	(0.0003)
Observations	43
R-squared	0.31

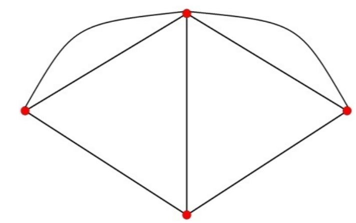
Regress MF
on

(Normalized) Centrality:	Eigen	Degree	Close	Bonacich	Btwn
	1.723*	**.177**	**.804**	**.024**	**.046**
	(.984)	**(.118)**	**(.481)**	**(.030)**	**(.032)**
Covariates:					
numHH					
shg					
savings					
fracGM					
Observations	**43**	**43**	**43**	**43**	**43**
R-squared	**.324**	**.314**	**.309**	**.278**	**.301**

2.5b: Application –
A new centrality measure: Diffusion Centrality



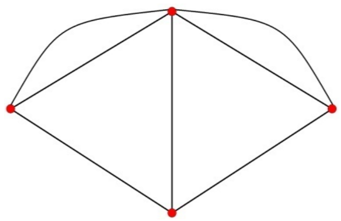
Diffusion Centrality: $DC_i\ (p,T)$



How many nodes are informed if:

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

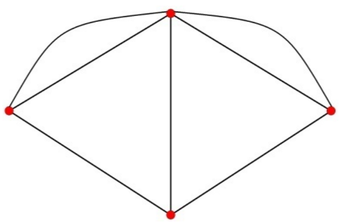
Diffusion Centrality



- DC $(p,T) = \sum_{t=1\ldots T} (p\mathbf{g})^t \mathbf{1}$

  If T=1: proportional to degree

Diffusion Centrality

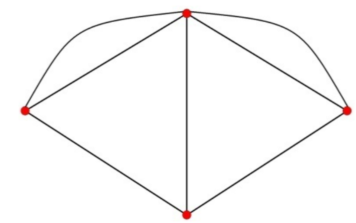


- DC $(p,T) = \sum_{t=1...T} (p\mathbf{g})^t \mathbf{1}$

If T=1: proportional to degree

If $p < 1/\lambda_1$ and T is large, becomes Katz-Bonacich

Diffusion Centrality



- DC (p,T) = $\Sigma_{t=1...T}\ (p\mathbf{g})^t\ \mathbf{1}$

If T=1: proportional to degree

If p< $1/\lambda_1$ and T is large, becomes Katz-Bonacich

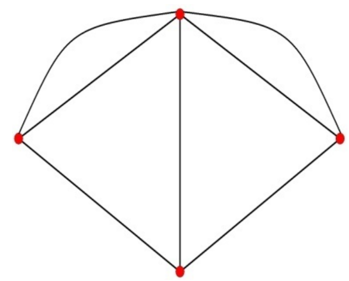
If p≥ $1/\lambda_1$ and T is large, becomes eigenvector

Regress
MF on

(Normalized)						
Centrality:	DC	Eigen	Degree	Close	Bonacich	Btwn
	.429*** **(.127)**	**1.723*** **(.984)**	**.177** **(.118)**	**.804** **(.481)**	**.024** **(.030)**	**.046** **(.032)**
Covariates: **numHH** **shg** **savings** **fracGM**						
Observations	**43**	**43**	**43**	**43**	**43**	**43**
R-squared	**.47**	**.324**	**.314**	**.309**	**.278**	**.301**

VARIABLES	mf	mf	mf
eigLeader		**2.22****	**1.07**
		(1.10)	**(0.89)**
diffuseCent.	**.54*****		**.49*****
	(0.15)		**(0.17)**
degreeLeader	**-.0006**	**-.004**	**-.0002**
	(.002)	**(.003)**	**(.0002)**
numHH	**-0.0004****	**-0.0002**	**-0.0002**
	(0.0002)	**(0.0002)**	**(0.0002)**
shg	**-.230**	**-.185**	**-.235**
	(.150)	**(.146)**	**(.138)**
savings	**-.337****	**-.149**	**-.321****
	(.144)	**(.114)**	**(.132)**
fracGM	**-.043**	**-.019**	**-.035**
	(.034)	**(.037)**	**(.036)**
Constant	**0.936*****	**0.461***	**.799*****
Observations	**43**	**43**	**43**
R-squared	**0.48**	**0.35**	**0.50**

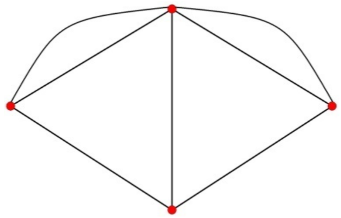
Social and Economic Networks: Models and Analysis



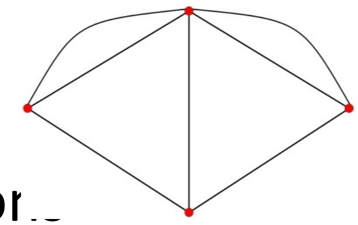
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.6: Random Networks

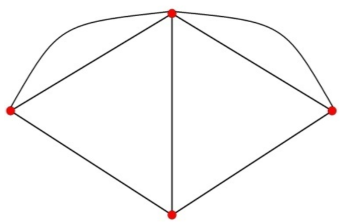


Summary so far:



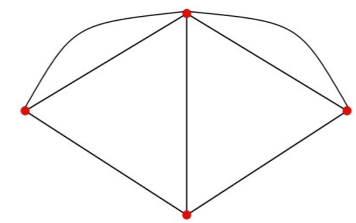
- Networks are prevalent and important in many interactions (labor markets, crime, garment industry, risk sharing...)
- Although complex, social networks have identifiable characteristics:
  - ``small'' average and maximum path length
  - high clustering relative to Poisson networks
  - degree distributions that exhibit different shapes
  - homophily – strong tendency to associate with own type
  - assortativity, strength of weak ties,...
  - a variety of centrality/influence/prestige measures...
- *Room for studies of methods...*

Outline



- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

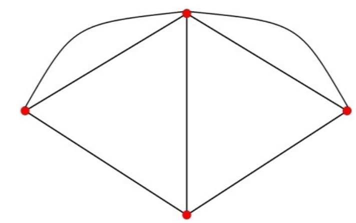
Questions



- Which networks form?
  - random graph models - ``How''
  - Economic/game theoretic models - ``Why''

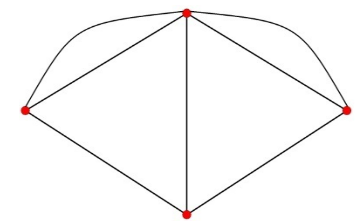
- How does it depend on context?

Static Random Networks



- Useful Benchmark
  - component structure
  - diameter
  - degree distribution
  - clustering...
- Tools and methods
  - properties and thresholds

E-R, Poisson Random Networks : G(n,p)

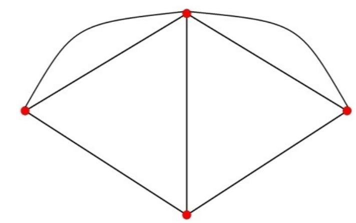


- independent probability p of each link
- probability that node has d links is **binomial**

  $[\ (n-1)!\ /\ (d!(n-d-1)!)\ ]\ p^d\ (1-p)^{n-d-1}$

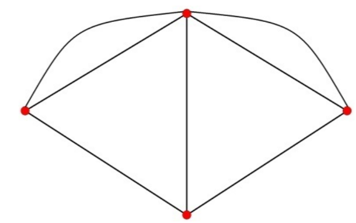
- Large n, small p, this is approximately a **Poisson** distribution: $[\ (n-1)^d\ /\ d!\ ]\ p^d\ e^{-(n-1)p}$

Properties of Networks



- Every network has some probability of forming
- How to make sense of that?
- Examine what happens for ``large'' networks
  - Bollobas (1985) book is a classic reference on random graph theory and many such results

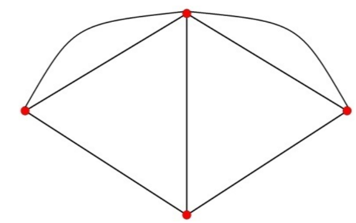
Specifying Properties



- G(N) = all the undirected networks on the set of nodes N

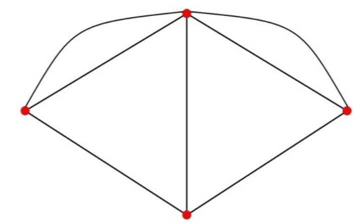
- A ***property*** is a set A(N) for each N such that A(N) is a subset of G(N)
  - a specification of which networks have that property

Examples of Properties



- A(N)={g | N_i (g) nonempty for all i inN}
  - property of no isolated nodes
- A(N)={g | ℓ (i,j) finite for all i,j in N}
  - network is connected
- A(N)={g | ℓ (i,j) < log(n) for all i,j inN}
  - diameter is less than log (n)

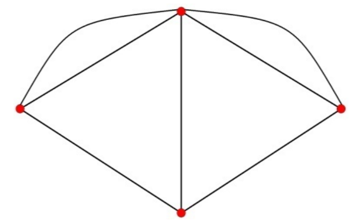
Monotone Properties



- A property A(N) is **monotone** if g in A(N) and g subset g’ implies g’ in A(N).

- All three of the previous properties are monotone

Limiting Properties

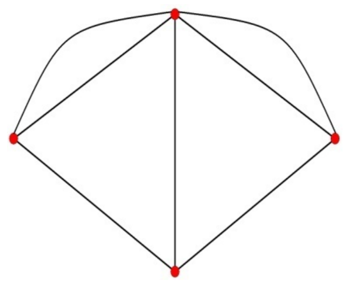


- In order to deduce things about random networks, we often look at ``large'' networks, by examining limits

- Examine a sequence of Erdos-Renyi Poisson random networks, with probability p(n)

- Deduce things about properties as $n \to \infty$

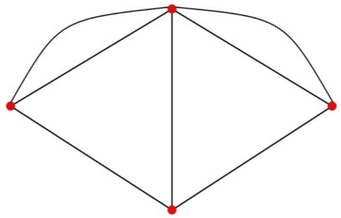
Social and Economic Networks: Models and Analysis



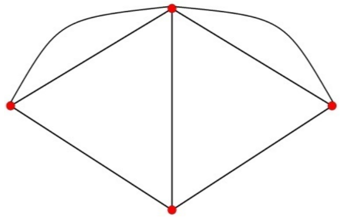
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.7: Random Networks Thresholds and Phase Transitions



Threshold Functions and Phase Transitions



- t(n) is a **threshold function** for a monotone property A(N) if

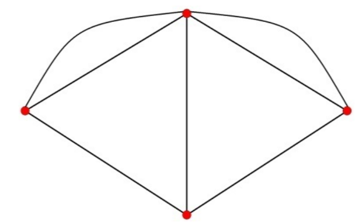
  Pr[A(N) | p(n)] -> 1 if p(n)/t(n) -> infinity

  and

  Pr[A(N) | p(n)] -> 0 if p(n)/t(n) ->0

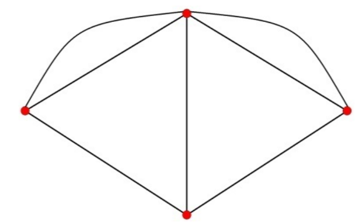
- A ***phase transition*** occurs at t(n)

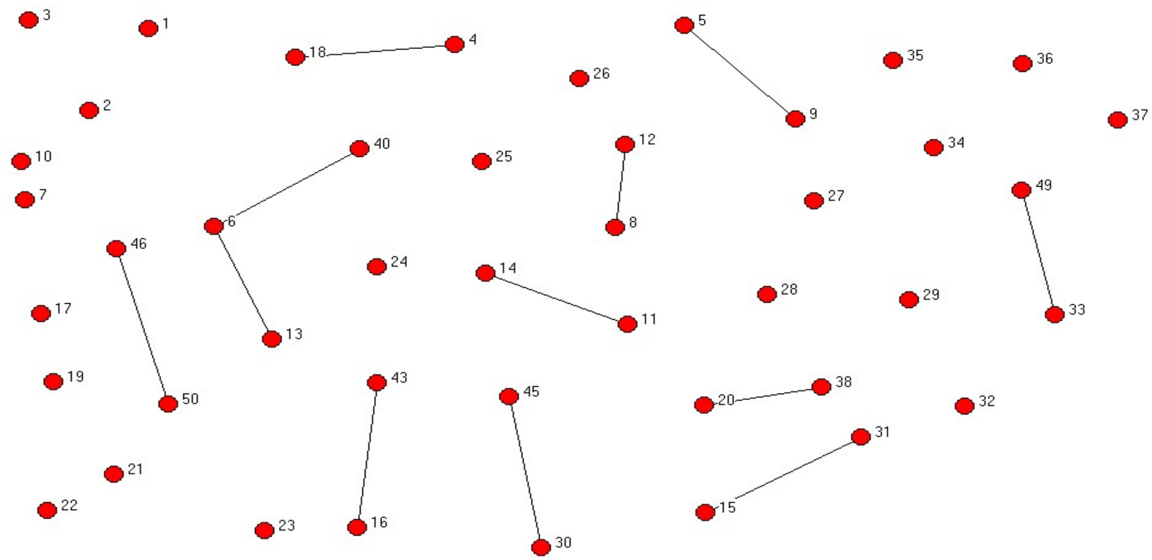
Thresholds for Poisson Random Networks:



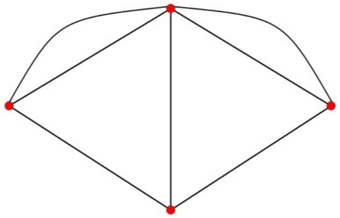
- $1/n^2$ - the network has some links (avg deg $1/n$)
- $1/n^{3/2}$ – the network has a component with at least three links (avg deg $1/n^{1/2}$)
- $1/n$ – the network has a cycle, the network has a unique giant component: a component with at least n^a nodes some fixed $a<1$; (avg deg 1)
- $\log(n)/n$ - the network is connected; (avg deg $\log(n)$)

Poisson p=.01, 50 nodes

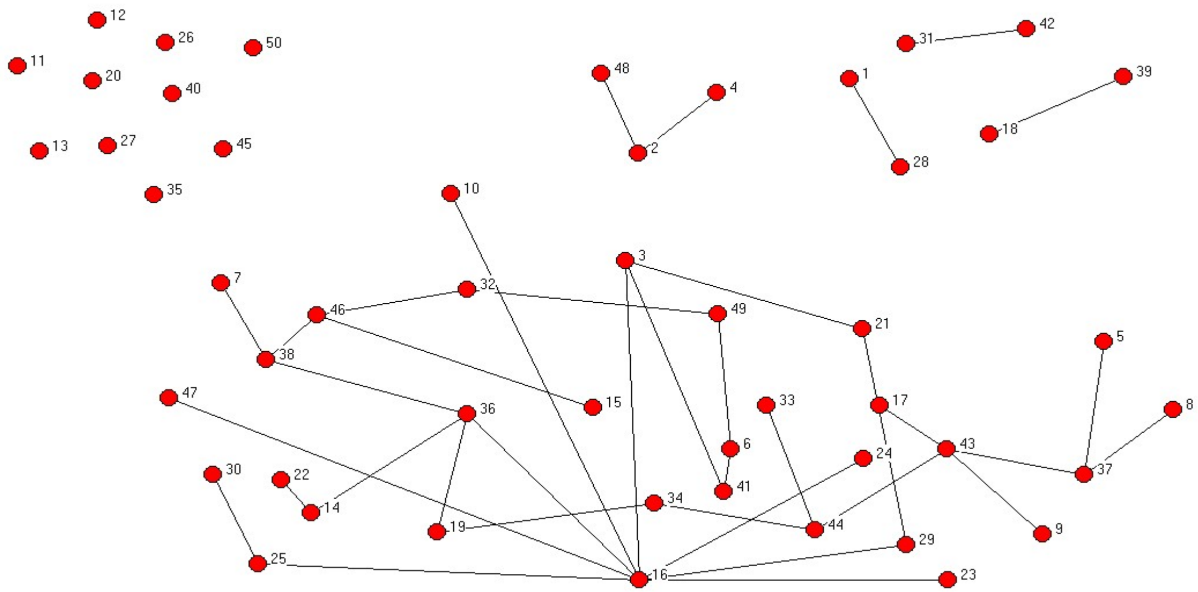




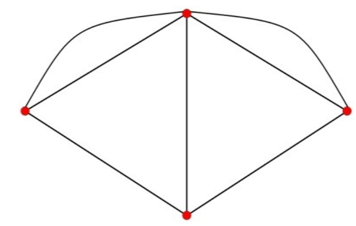
Poisson p=.03, 50 nodes

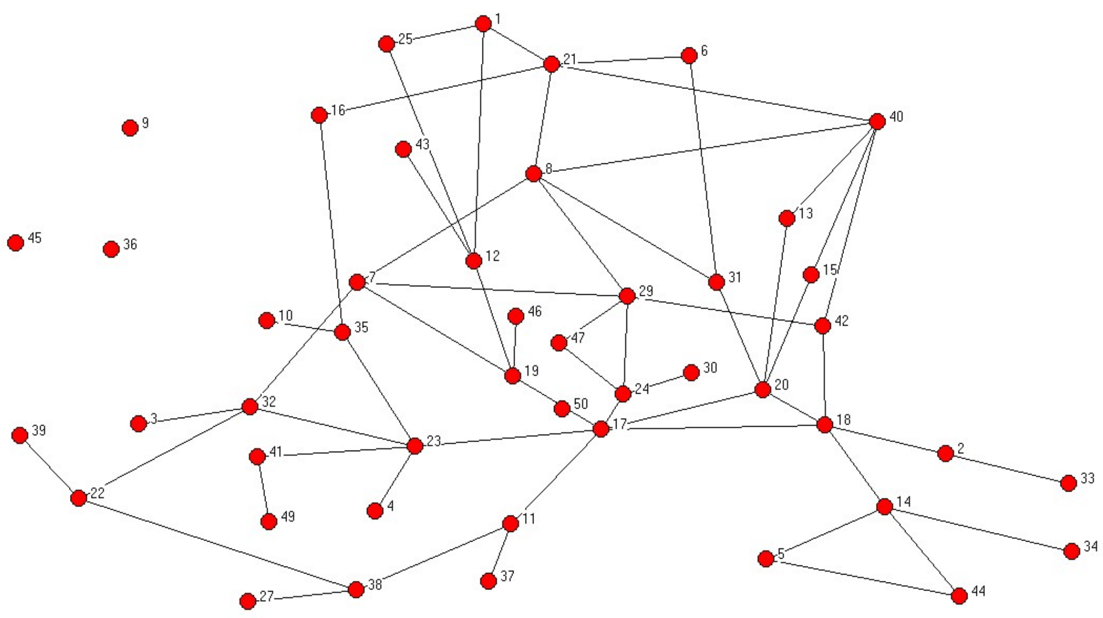


.02 is the threshold for emergence of cycles and a giant component

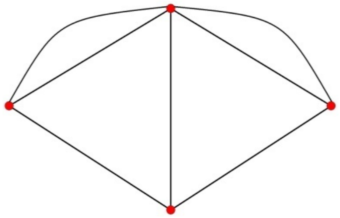


Poisson p=.05, 50 nodes

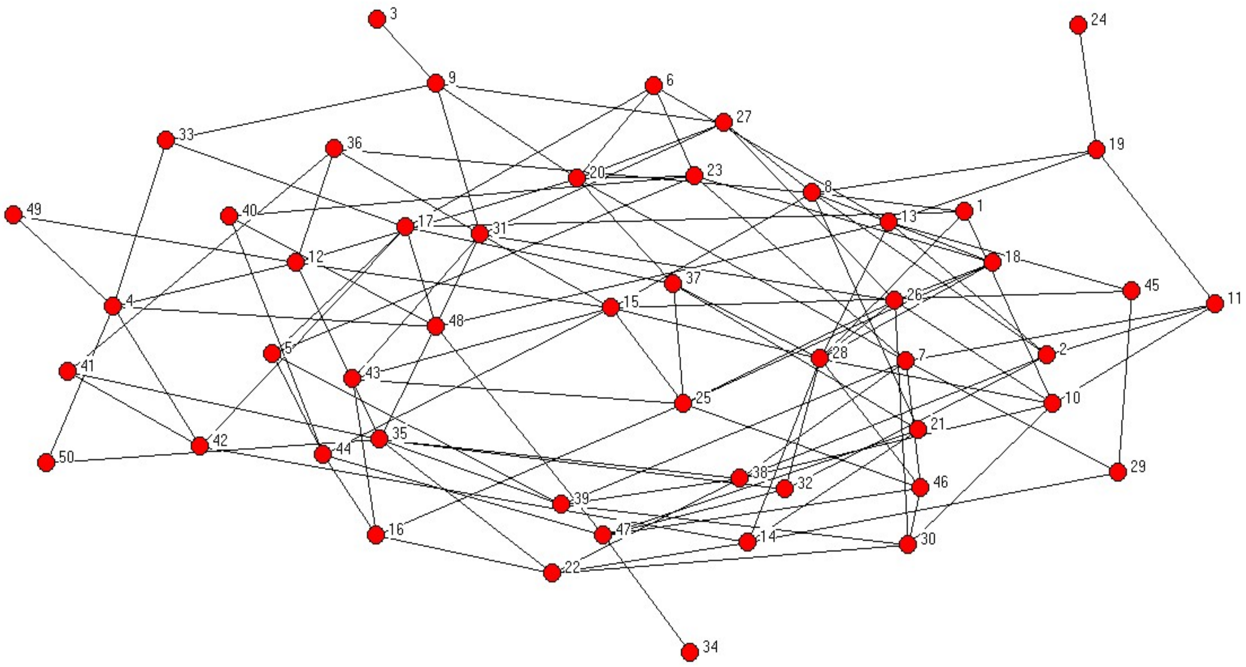




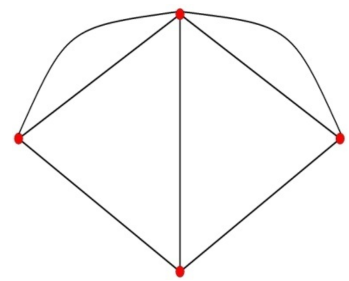
Poisson p=.10, 50 nodes



.08 is the threshold for connection



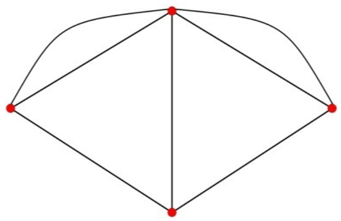
Social and Economic Networks: Models and Analysis



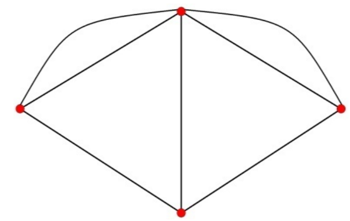
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.8: A Threshold Theorem

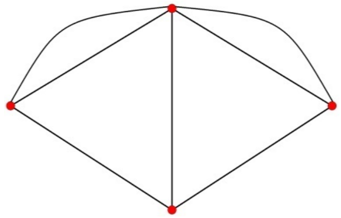


A Threshold Theorem:



Theorem [Erdos and Renyi 1959] A threshold function for the connectedness of a Poisson random network is $t(n)=\log(n)/n$

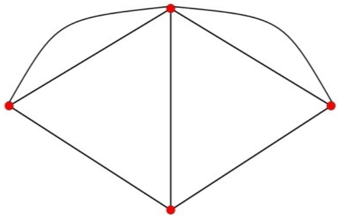
Part of the Proof:



1. Show that if p(n)/t(n) -> 0 then there will be isolated nodes with probability 1.
2. Show that if p(n)/t(n) -> infinity then there will not be any components of size less than n/2 with probability 1.

Show 1 – intuition for rest is that threshold for isolated node is the same as threshold for small component

Useful Approximations



Definition of exponential function:

$$e^x = \lim_n \ (1+ x/n)^n$$

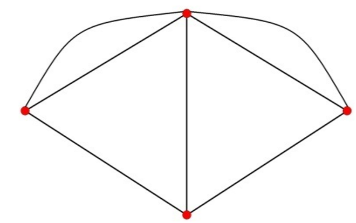
Taylor series approximation:

$$e^x = 1 + x + x^2/2! + x^3/3! \ldots$$

$$= \sum \ x^n / n!$$

$$[\ f(x) = f(a) + f'(a)(x-a)/1! + f''(a)(x-a)^2/2! \ldots. \ ]$$

Let us examine the logic

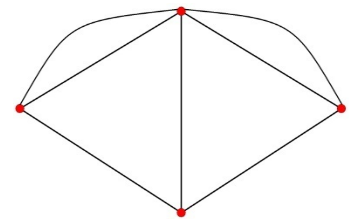


Let us show that E[d]=log(n) is the threshold above which we expect each node to have some links

In fact, above this threshold we expect each node to have many links

Once every node has many links, the chance of disconnected components vanishes

E[d] = log(n) is ``isolates " threshold:



- Rewrite $E[d] = p(n-1) = r + \log(n)$ for some r

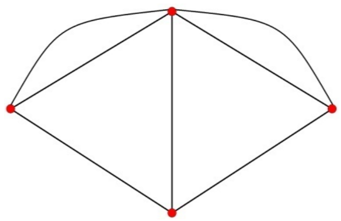
- Probability that some node is isolated is probability that it has no links

- Probability that some link is not present is (1-p)

- Links are independent, so probability of isolation is

$$(1-p)^{n-1}$$

E[d]=log(n) is isolates threshold:



- Probability that some node is isolated is

$$(1-p)^{n-1} = (1-(r + \log(n))/(n-1))^{n-1}$$

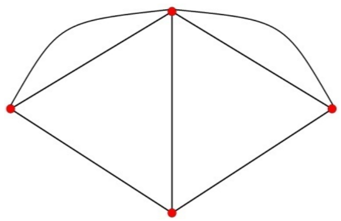
- Recall that $(1-x/n)^n$ approaches e^{-x}

(if x/n vanishes - so let us consider that case - other cases are more extreme and so easy to fill in the missing steps...)

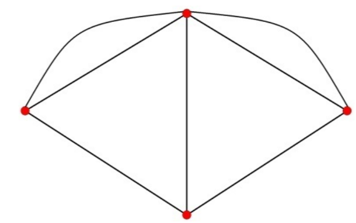
- Probability that some node is isolated is

$$(1- (r + \log(n))/(n-1))^{n-1} = e^{-r - \log(n)} = e^{-r}/n$$

E[d]=log(n) is isolates threshold:

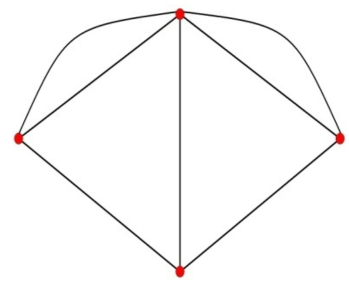


- Expected number of isolated nodes is e^{-r}
- $E(d) - \log(n) = r \to \infty$ implies Expected number of isolated nodes goes to 0
- $E(d) - \log(n) = r \to -\infty$ implies that expected number of isolated nodes becomes infinite.
  [E.g., E(d) bounded by M implies $r \to -\log n + M$
  number of expected isolated nodes goes to $n\, e^{-M}$



- So, the expected number of isolated nodes $= e^{-r(n)}$ goes to 0 if r(n) tends to infinity and to infinity if r(n) tends to minus infinity.
- If the expected number tends to 0 then the probability of having one tends to 0
- If the expected number tends to infinity, then extra step using Chebyshev and showing that the variance is no more than twice the mean shows the probability of having one goes to 1.

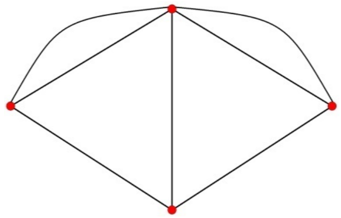
Social and Economic Networks: Models and Analysis



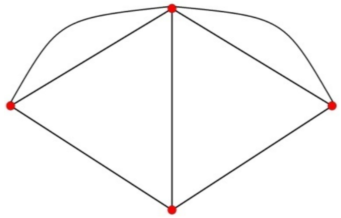
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

2.9: A Small World Model

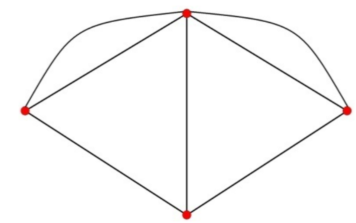


Outline



- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

Other Static Models:

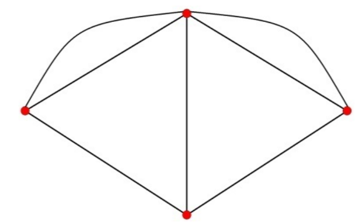


- Models to generate clustering

- Models to generate other than Poisson degree distributions

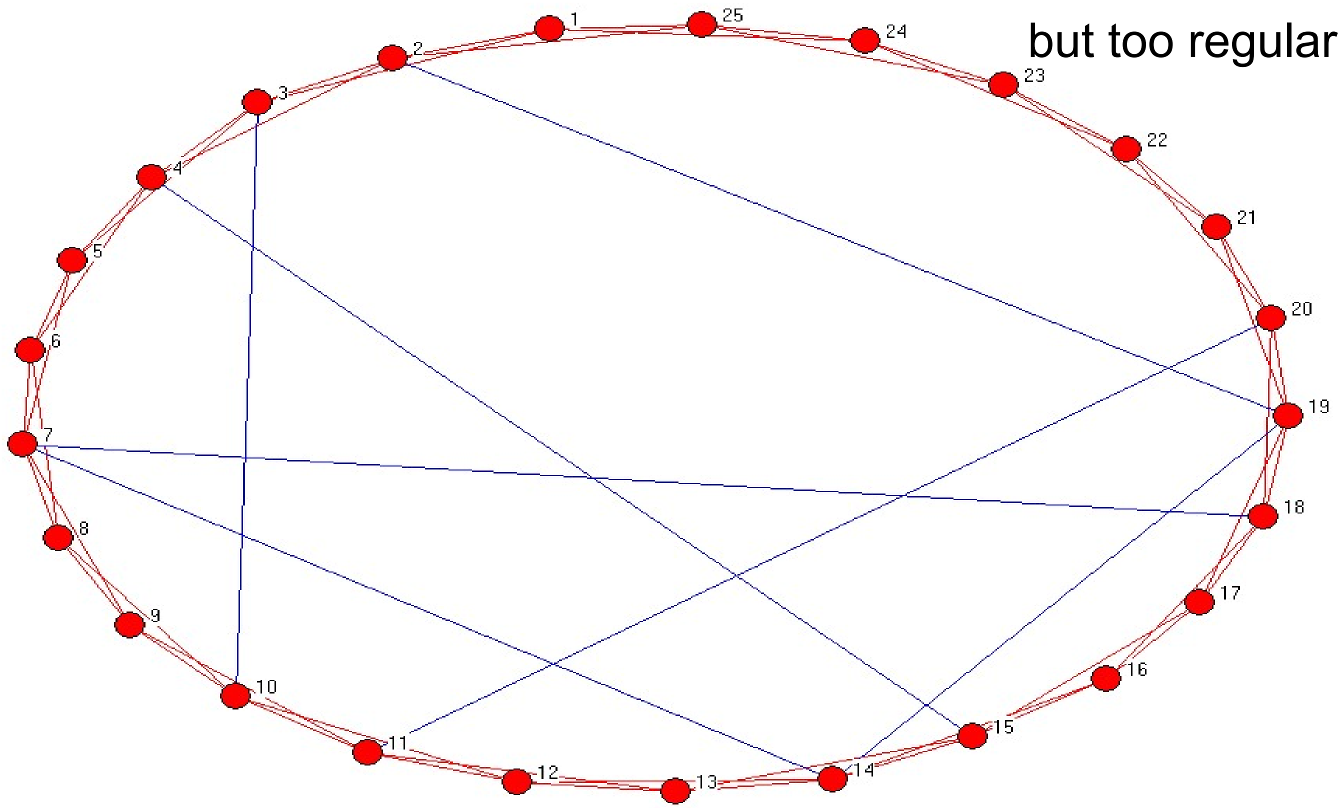
- Models to fit to data

Rewired lattice -Watts and Strogatz 98



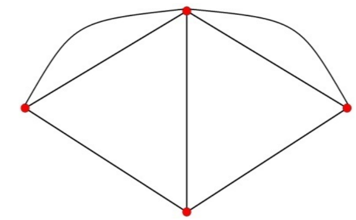
- Erdos-Renyi model misses clustering
  - clustering is on the order of p; going to 0 unless average degree is becoming infinite (and highly so...)
- Start with ring-lattice and then randomly pick some links to rewire
  - start with high clustering but high diameter
  - as rewire enough links, get low diameter
  - don’t rewire too many, keep high clustering

Rewired lattice example



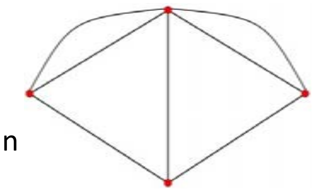
high clustering
low diameter if degree is high
but too regular

Week 2 Wrap



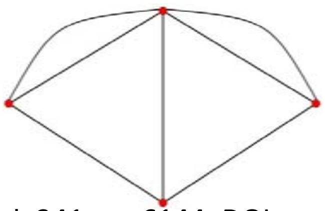
- Networks based on characteristics: homophily
- Local aspects, positions: centrality measures
- Random networks: sharp thresholds, properties, phase transitions
- Small worlds: combining few random links gives tree-like structure necessary to shorten paths without destroying local clustering

Week 2: References in Order Mentioned



- Lazarsfeld, P.F., and R.K. Merton (1954) “Friendship as a Social Process: A Substantive and Methodological Analysis,” in *Control in Modern Society,* M. Berger, ed., New York: Van Nostrand.*ociology* 98:1259–1319.
- Shrum,W., N.H. Cheek, Jr., and S.M. Hunter (1988) “Friendship in School: Gender and Racial Homophily,” *Sociology of Education 61:227–239.*
- Blau, P.M. (1964) *Exchange and Power in Social Life, New York: Wiley.*
- —(1977) *Inequality and Heterogeneity: A Primitive Theory of Social Structure,* New York: Free Press.
- Marsden, P.V. (1987) “Core Discussion Networks of Americans,” *American Sociological Review 52:122–131.*
- — (1988) “Homogeneity in Confiding Relations,” *Social Networks 10:57–76.*
- Moody, J. (2001) “Race, School Integration, and Friendship Segregation in America,” *American Journal of Sociology 107(3):679–716.*
- McPherson, M., L. Smith-Lovin, and J.M. Cook (2001) “Birds of a Feather: Homophily in Social Networks,” *Annual Review of Sociololgy 27:415–444*
- Fryer, R. (2007) “Guess Who’s Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century,” *Journal of Economic Perspectives 21(2):71–90.*
- Verbrugge, L.M (1977) The Structure of Adult Friendship Choices , Social forces, 56:2, 576-597
- Currarini S, Jackson M.O., Pin P. (2009) “An economic model of friendship: homophily, minorities and segregation,” Econometrica, Vol. 77, No. 4, 1003-1045,
- Currarini, S., M.O. Jackson, and P. Pin (2010) “Identifying the roles of race-based choice and chance in high school friendship network formation,” in the Proceedings of the National Academy of Sciences, 107(11): 4857 – 4861
- Jackson M.O. 2007 Social Structure, Segregation, and Economic Behavior, Nancy Schwartz Memorial Lecture, given in April 2007 at Northwestern University, printed version: http://www.stanford.edu/jacksonm/schwartzlecture.pdf
- Baerveldt, C., M.A.J. Van Duijn, L. Vermeij, and D.A. Van Hemert (2004) “Ethnic Boundaries and Personal Choice: Assessing the Influence of Individual Inclinations to Choose Intra-Ethnic Relationships on Pupils’ Networks,” *Social Networks 26:55*– 74.

Week 2: References Cont'd



- Granovetter M. 1973. The strength of weak ties. Am. J. Sociol. 78: 1360 – 80
- `Banerjee, A., A.G. Chandrasekhar, E.Duflo, M.O. Jackson D ``The Diffusion of Microfinance'' *Science*, 26 July 2013: Vol. 341 no. 6144, DOI: 10.1126/science.1236498.
- Bollob´as, B. (1985 -2001) *Random Graphs, second edition, Cambridge: Cambridge University* Press.
- Watts, D.J., and S. Strogatz (1998) "Collective Dynamics of 'Small-World' Networks," *Nature 393:440–442.*