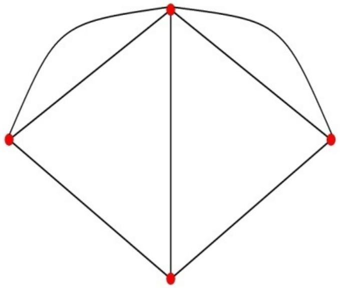
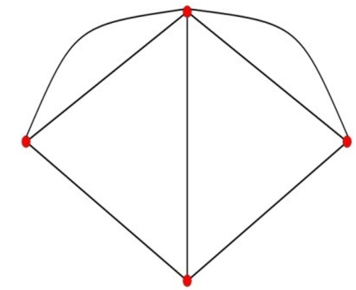


Social and Economic Networks

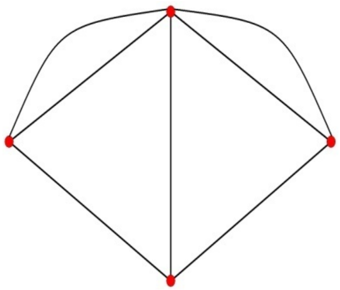
Matthew O. Jackson

Network Formation



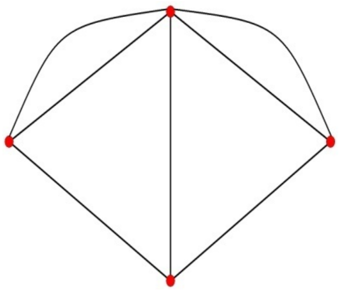
- Which networks form?
  - random graph models -- ``How’’
  - Economic/game theoretic models -- ``Why’’
- When do efficient networks form?
- How is formation affected by bargaining?
- How does it depend on context?

Uses of models



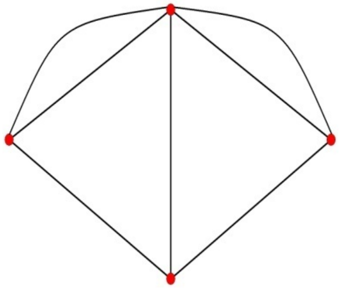
- Hypothesis testing
  - why clustering
  - why homophily...
- Counterfactuals, policy evaluation
- As an input into studying behavior on networks
  - Networks are endogenous!

Approaches



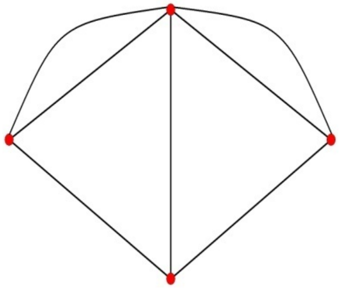
- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Approaches



- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Erdos-Renyi/Poisson Random Networks



- independent probability p of each link

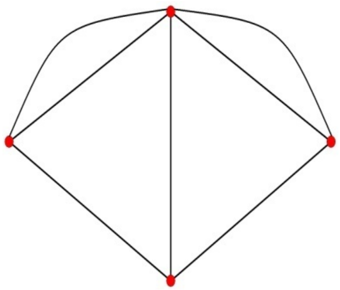
- probability that node has d links is **binomial**

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

- Large n, small p, this is approximately a **Poisson** distribution:

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

Threshold Functions and Phase Transitions



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

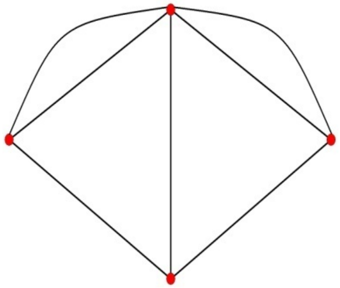
  Pr[A(N) | p(n)] -> 1 if p(n)/t(n) -> infinity

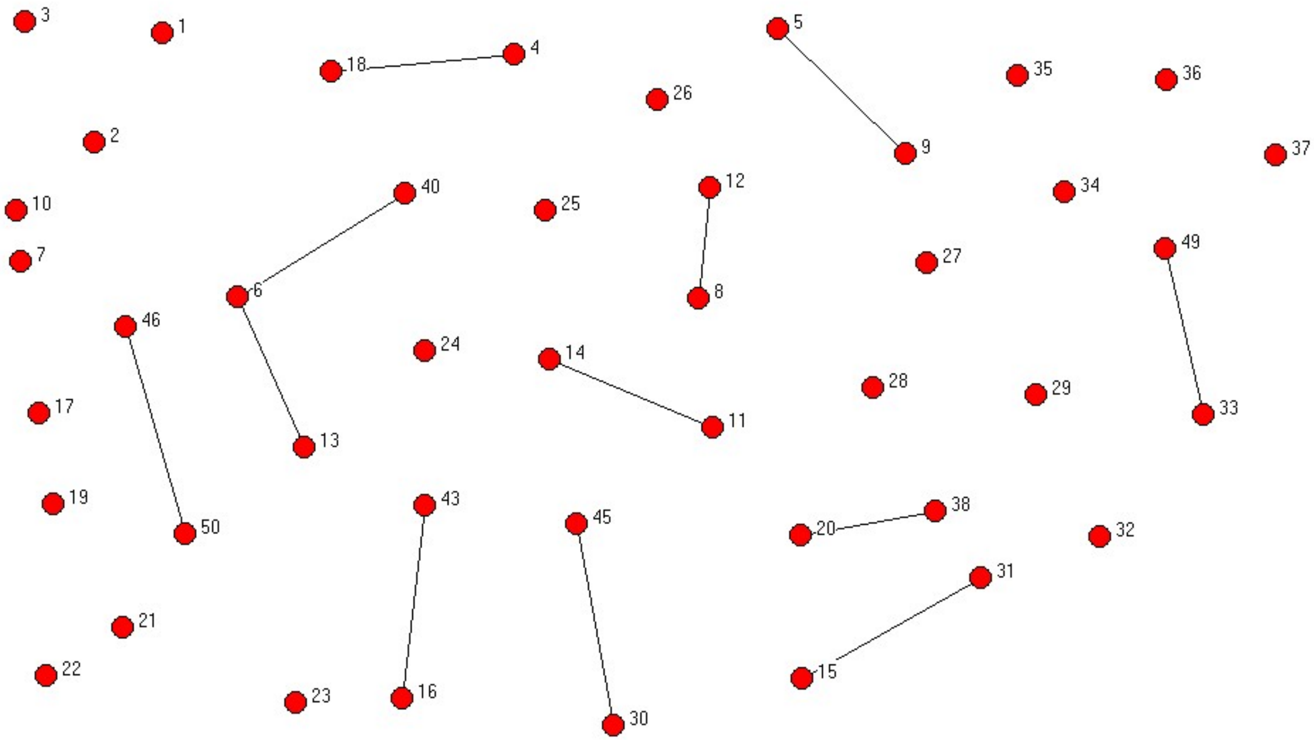
  and

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

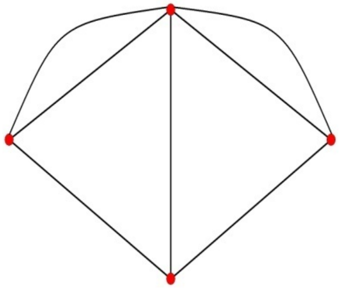
- A ***phase transition*** occurs at t(n)

Poisson p=.01, 50 nodes

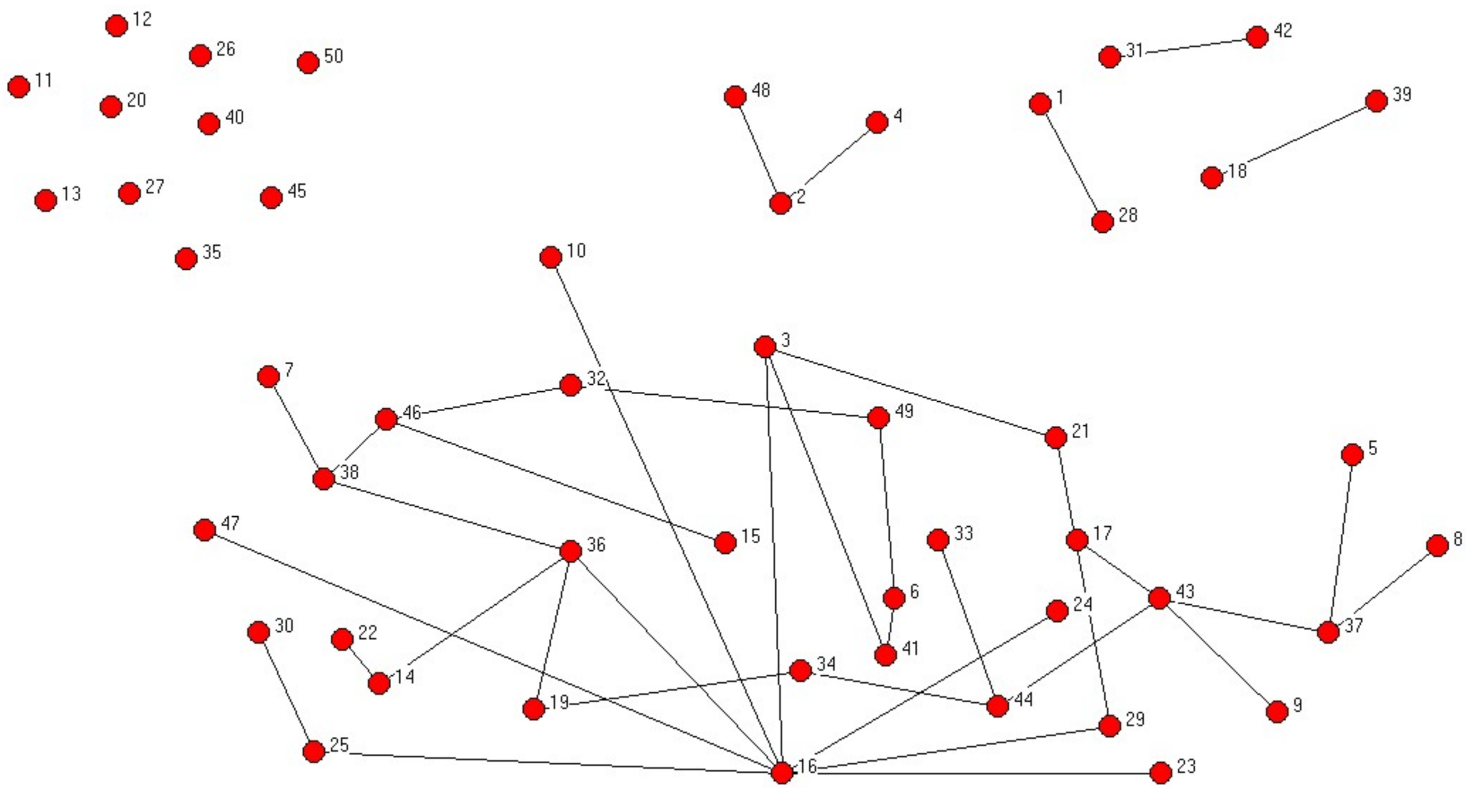




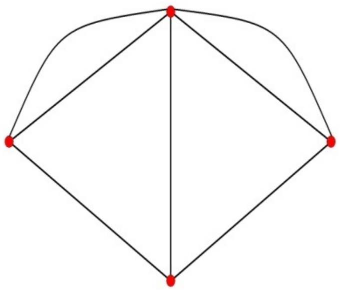
Poisson p=.03, 50 nodes

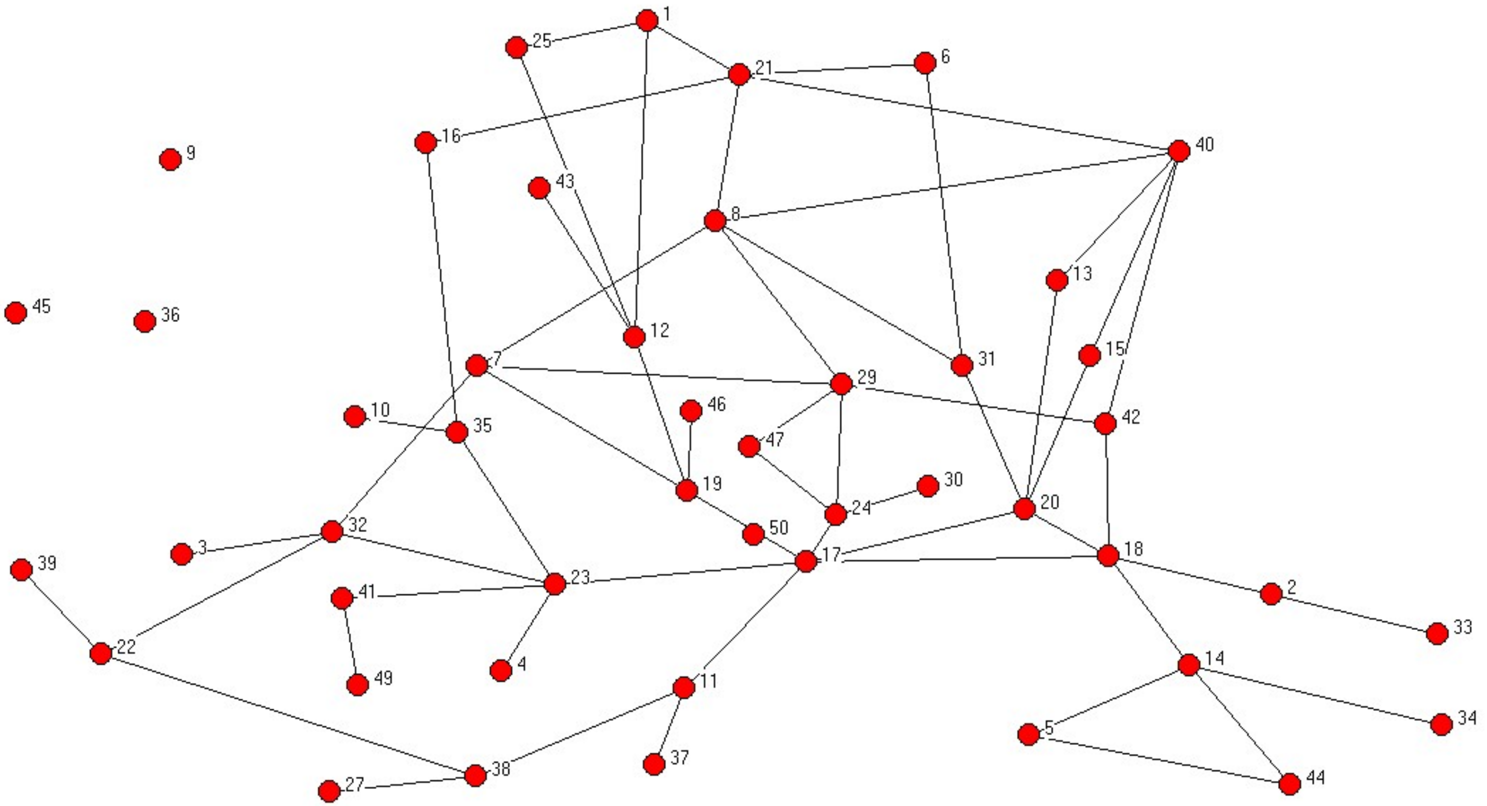


.02 is the threshold for emergence of cycles and a giant component

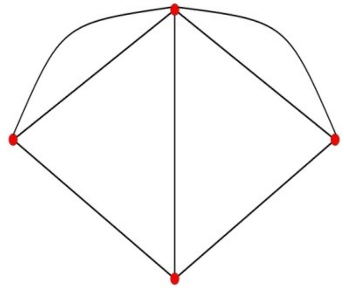


Poisson p=.05, 50 nodes

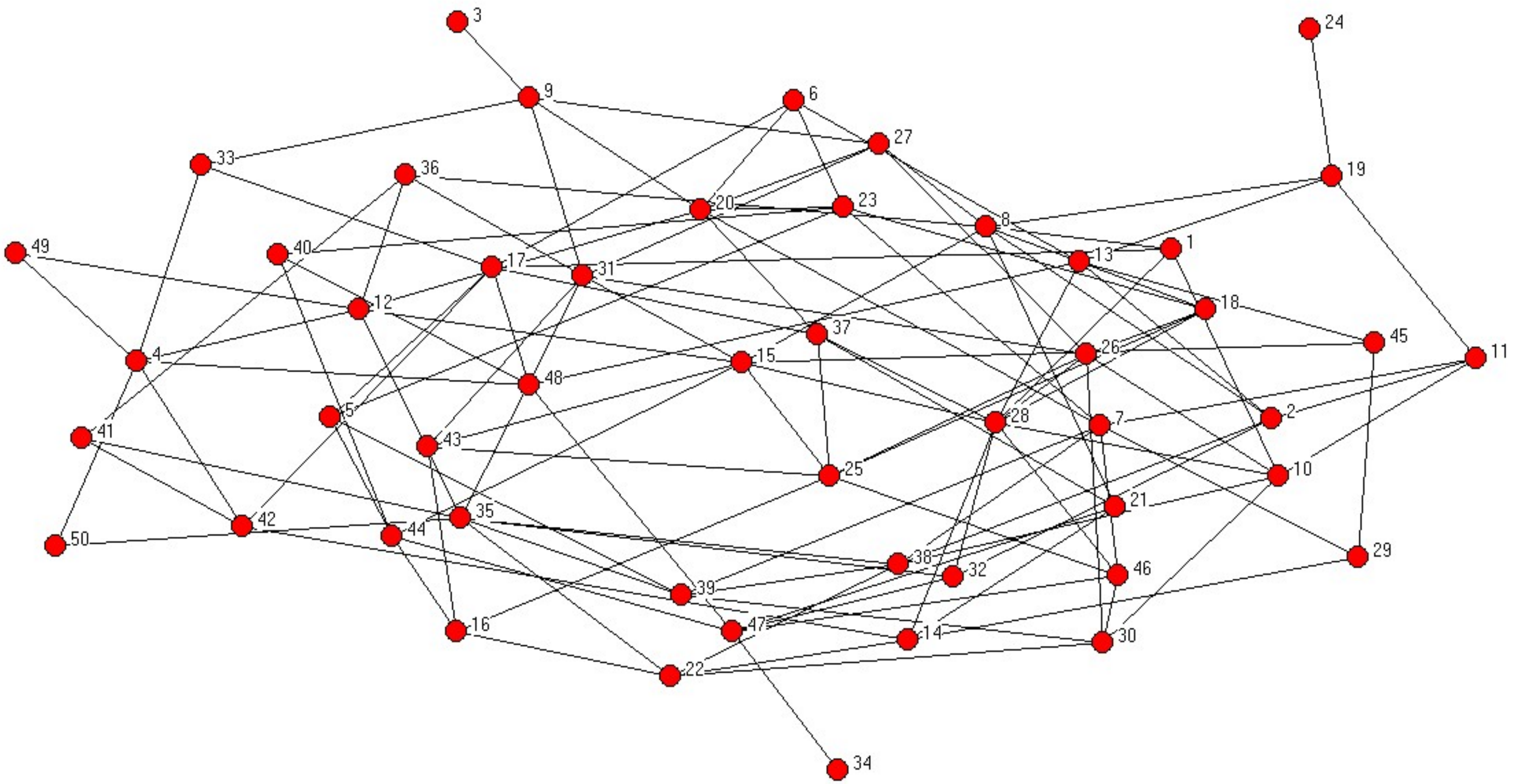




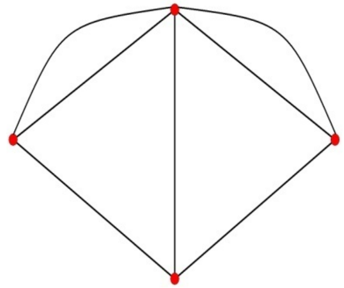
Poisson p=.10, 50 nodes



.08 is the threshold for connection

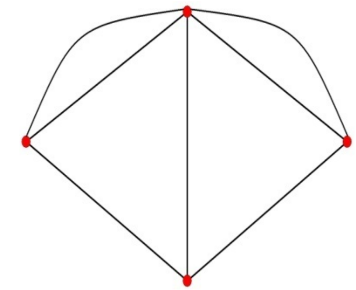


Thresholds for Poisson Random Networks:



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

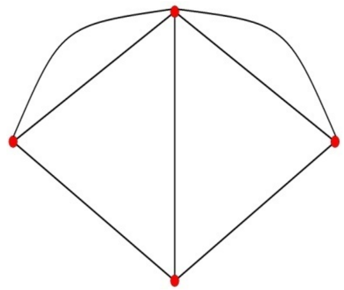
Rewired lattice -Watts and Strogatz (1998)



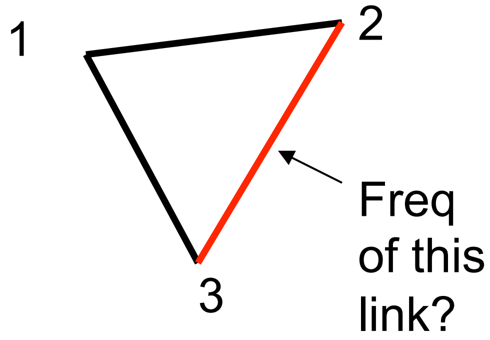
- Erdos-Renyi model misses clustering
  - clustering is on the order of p; going to 0 unless average degree is becoming infinite (and highly so...)

- 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

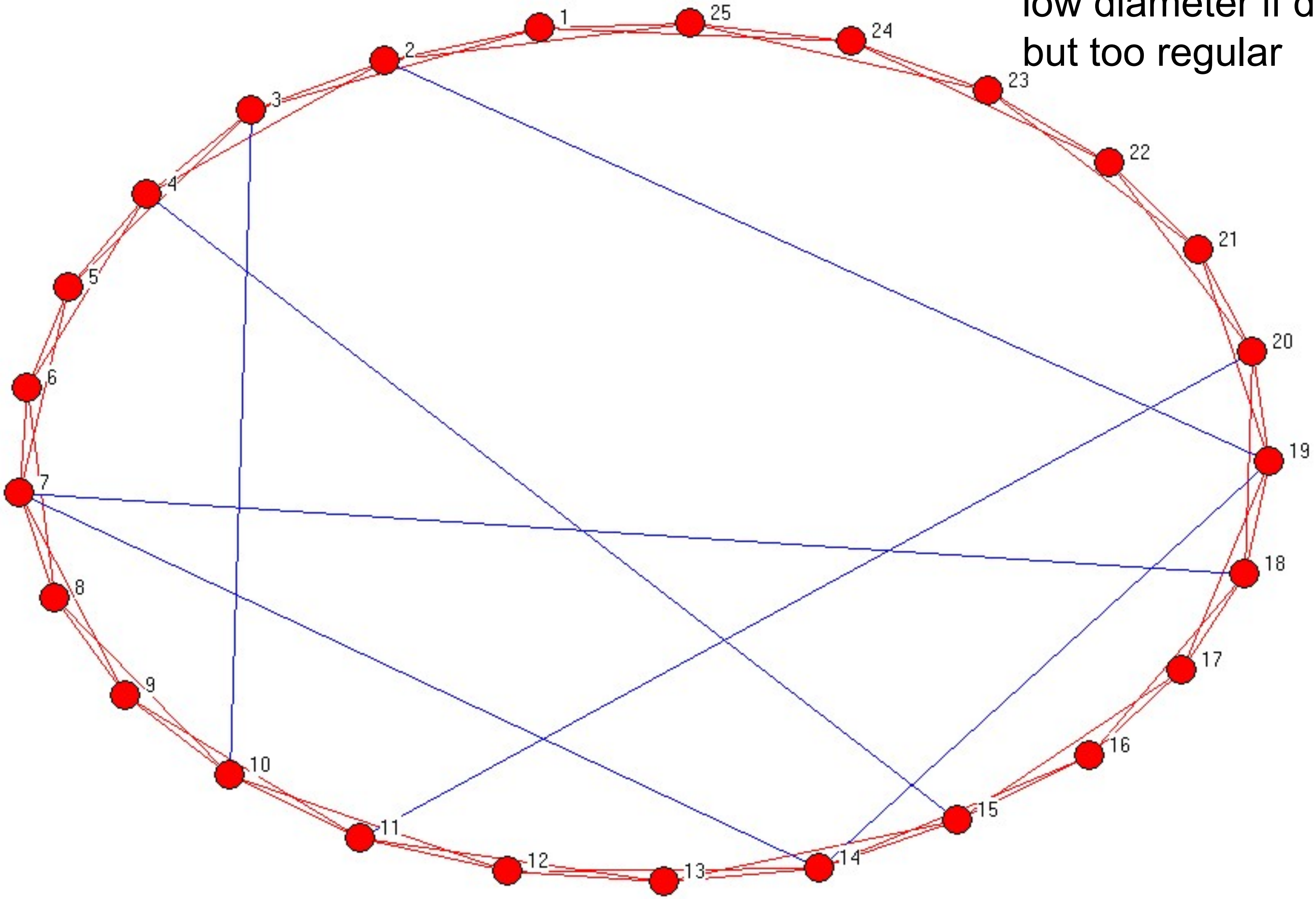
Clustering Coefficients



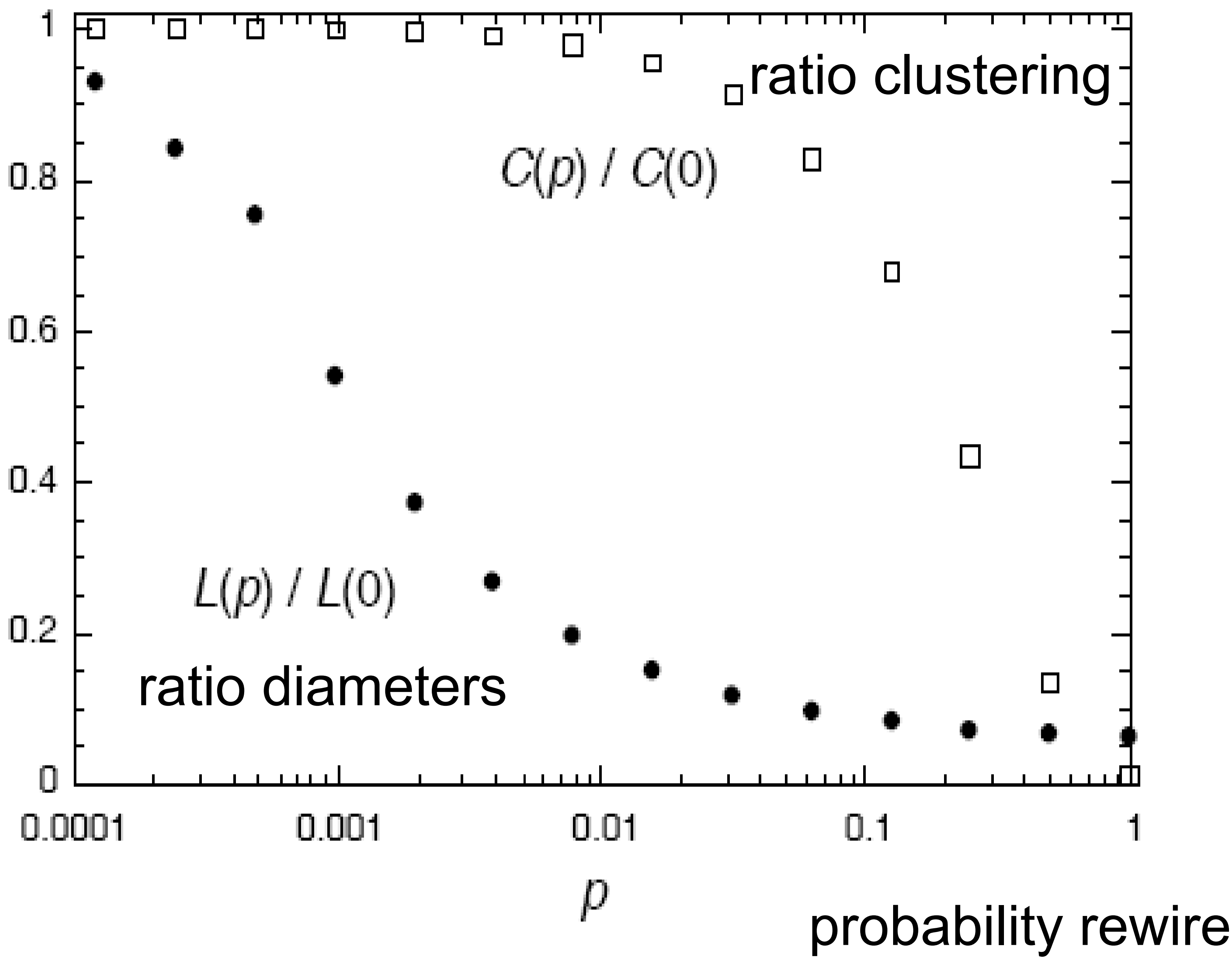
- Prison friendships
  - **.31** (MacRae 60) vs .0134
- co-authorships
  - **.15** math (Grossman 02) vs .00002,
  - **.09** biology (Newman 01) vs .00001,
  - **.19** econ (Goyal, van der Leij, Moraga 06) vs .00002,
- Florentine Marriage and Business dealings
  - **.46** on 15 central families vs .29...
- Web
  - **.11** for web (Adamic 99) vs .



Rewired lattice example

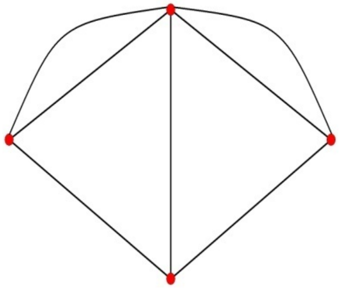


high clustering
low diameter if degree is high
but too regular



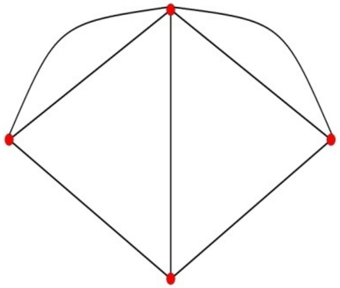
ratio clustering
C(p) / C(0)
L(p) / L(0)
ratio diameters
0
0.2
0.4
0.6
0.8
1
0.0001
0.001
0.01
0.1
1
p
probability rewire

Approaches



- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Economic Game Theoretic Models of Network Formation

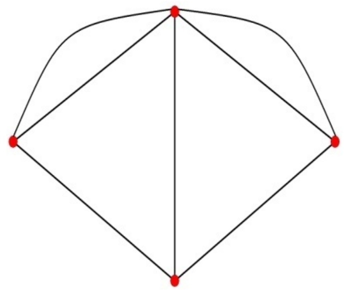


- Costs and benefits for each agent associated with each network

- Agents choose links

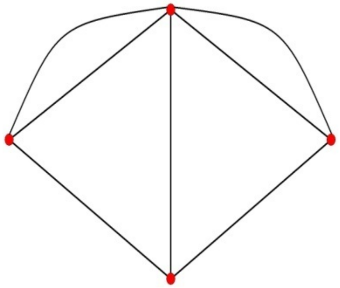
- Contrast incentives and social efficiency

Modeling Choices



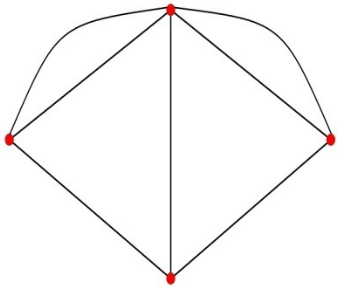
- How should we model incentives to form and sever links?
  - is consensus needed (undirected/directed)?
  - can they coordinate changes in the network?
  - is the process dynamic or static?
  - how sophisticated are agents?
  - what do they know when making a decision?
  - do they make errors?
  - what happens on the network?
  - can they compensate each other for relationship?
  - are links ajustable in intensity?

Some Questions



- Which networks are likely to form?
- Are some more stable than others to various perturbations?
- Are the networks that form efficient?
- How inefficient are they if they are not efficient?
- Can intervention help improve efficiency?
- Can such models provide insight into observed characteristics of networks?

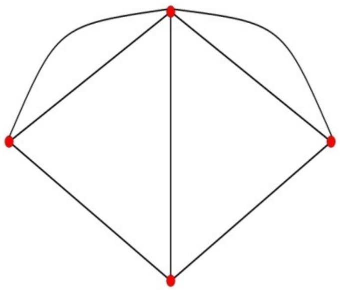
An Economic Analysis: Jackson Wolinsky (1996)



- u_i (g) - payoff to i if the network is g

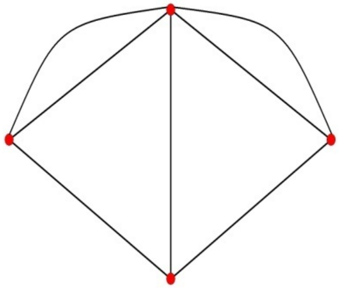
- undirected network formation

Connections Model JW96



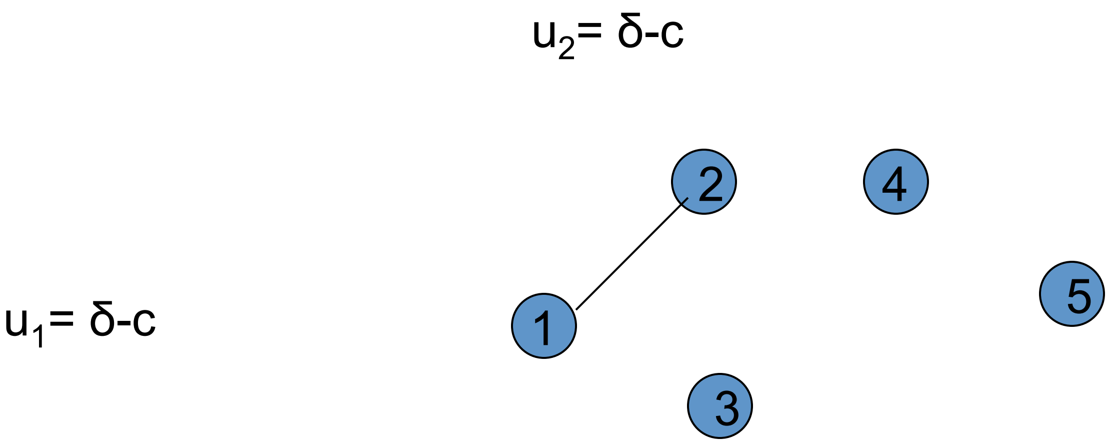
- $0 \leq \delta \leq 1$ a benefit parameter for i from connection between i and j
- $0 \leq c_{ij}$ cost to i of link to j
- $\ell(i,j)$ shortest path length between i,j

$$u_i(g) = \sum_j \delta^{\ell(i,j)} - \sum_{j \text{ in } N_i(g)} c_{ij}$$

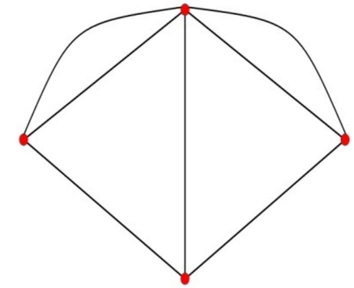


Symmetric Version:

- benefit from a friend is $\delta<1$
- benefit from a friend of a friend is δ^2,...
- cost of a link is $c>0$



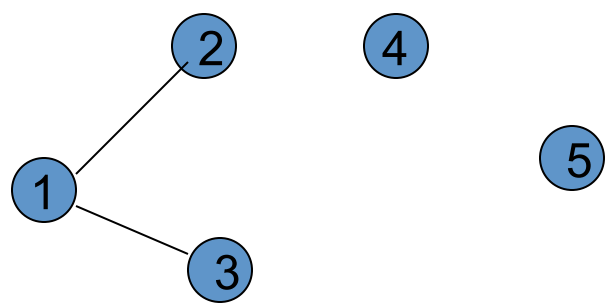
Symmetric Version:



- benefit from a friend is $\delta<1$
- benefit from a friend of a friend is δ^2,...
- cost of a link is $c>0$

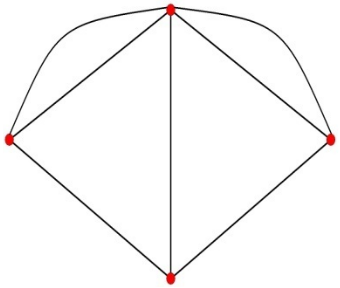
$u_2 = \delta + \delta^2 - c$

$u_1 = 2\delta - 2c$

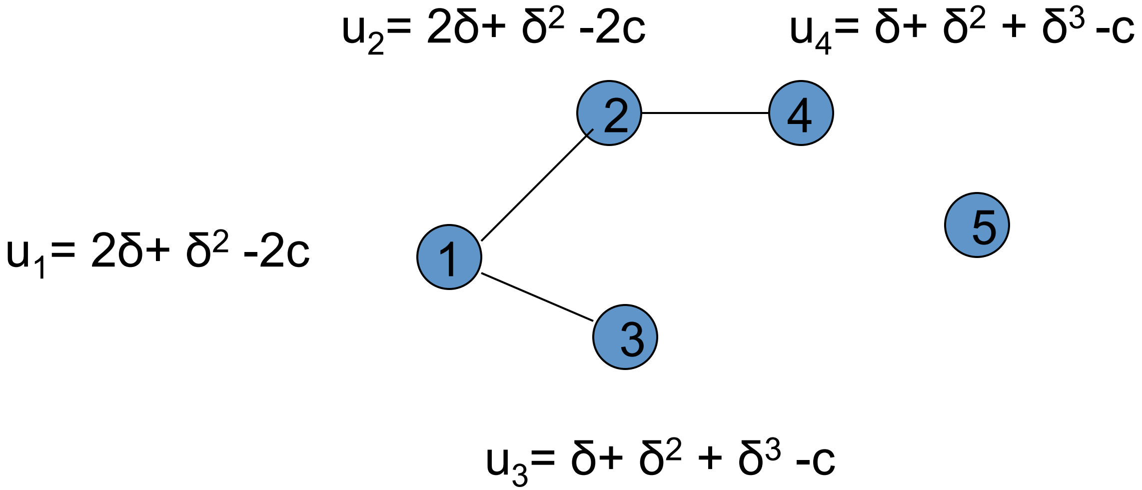


$u_3 = \delta + \delta^2 - c$

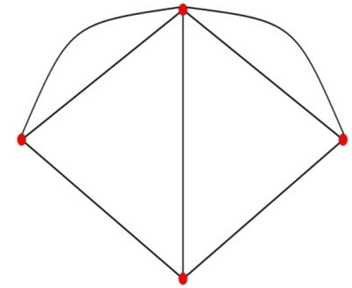
Symmetric Version:



- benefit from a friend is δ<1
- benefit from a friend of a friend is δ^2,…
- cost of a link is c>0



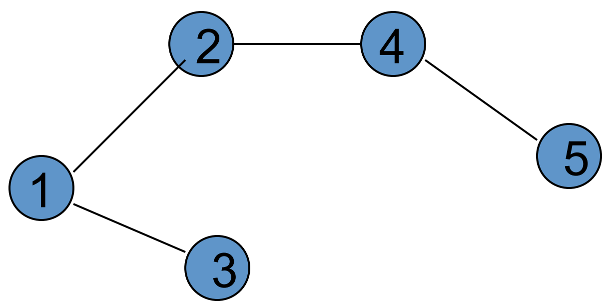
Symmetric Version:



- benefit from a friend is $\delta<1$
- benefit from a friend of a friend is δ^2,...
- cost of a link is $c>0$

$u_2 = 2\delta + 2\delta^2 - 2c$ $u_4 = 2\delta + \delta^2 + \delta^3 - 2c$

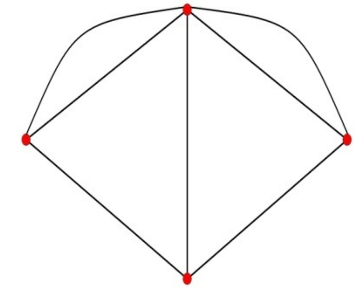
$u_1 = 2\delta + \delta^2 + \delta^3 - 2c$



$u_5 = \delta + \delta^2 + \delta^3 + \delta^4 - c$

$u_3 = \delta + \delta^2 + \delta^3 + \delta^4 - c$

Symmetric Version:

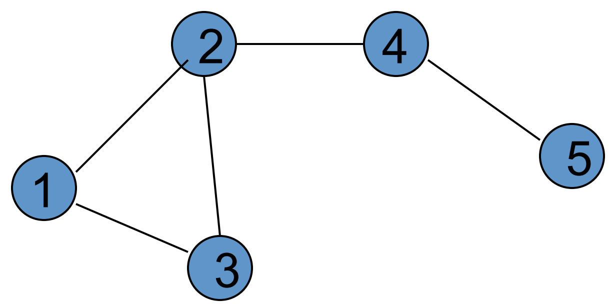


- benefit from a friend is $\delta<1$
- benefit from a friend of a friend is δ^2,...
- cost of a link is $c>0$

$u_2 = 3\delta + \delta^2 - 3c$

$u_4 = 2\delta + 2\delta^2 - 2c$

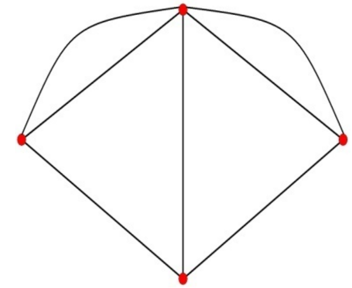
$u_1 = 2\delta + \delta^2 + \delta^3 - 2c$



$u_5 = \delta + \delta^2 + 2\delta^3 - c$

$u_3 = 2\delta + \delta^2 + \delta^3 - 2c$

Symmetric Version:

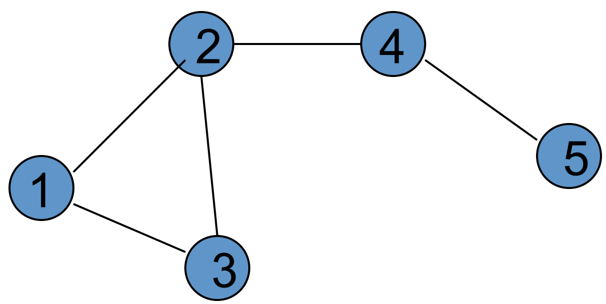


- benefit from a friend is $\delta<1$
- benefit from a friend of a friend is δ^2,...
- cost of a link is $c>0$

$u_2 = 3\delta + \delta^2 - 3c$

$u_4 = 2\delta + 2\delta^2 - 2c$

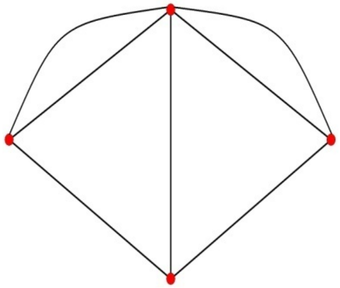
$u_1 = 2\delta + \delta^2 + \delta^3 - 2c$



$u_5 = \delta + \delta^2 + 2\delta^3 - c$

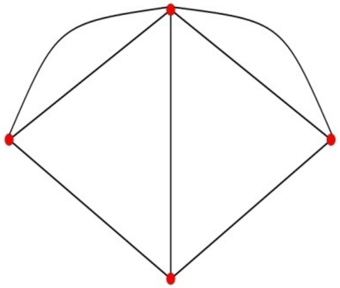
$u_3 = 2\delta + \delta^2 + \delta^3 - 2c$

Questions:



- Which network are best for society?
- Which networks are formed by the agents?

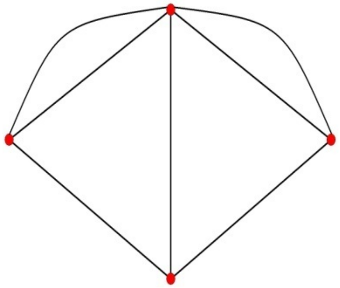
Modeling Incentives: Pairwise Stability



- no agent gains from severing a link – relationships must be beneficial to be maintained

- no two agents both gain from adding a link (at least one strictly) – beneficial relationships are pursued when available

Pairwise Stability



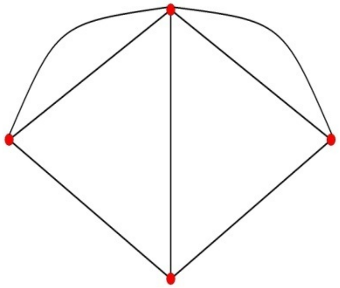
- $u_i(g) \geq u_i(g\text{-}ij)$ for i and ij in g
  - no agent gains from severing a link

- $u_i(g+ij) > u_i(g)$ implies $u_j(g+ij) < u_j(g)$ for ij not in g
  - no two agents both gain from adding a link (at least one strictly)

- a weak concept, but often narrows things down

Both are Nash equilibria, but only the dyad is pairwise stable

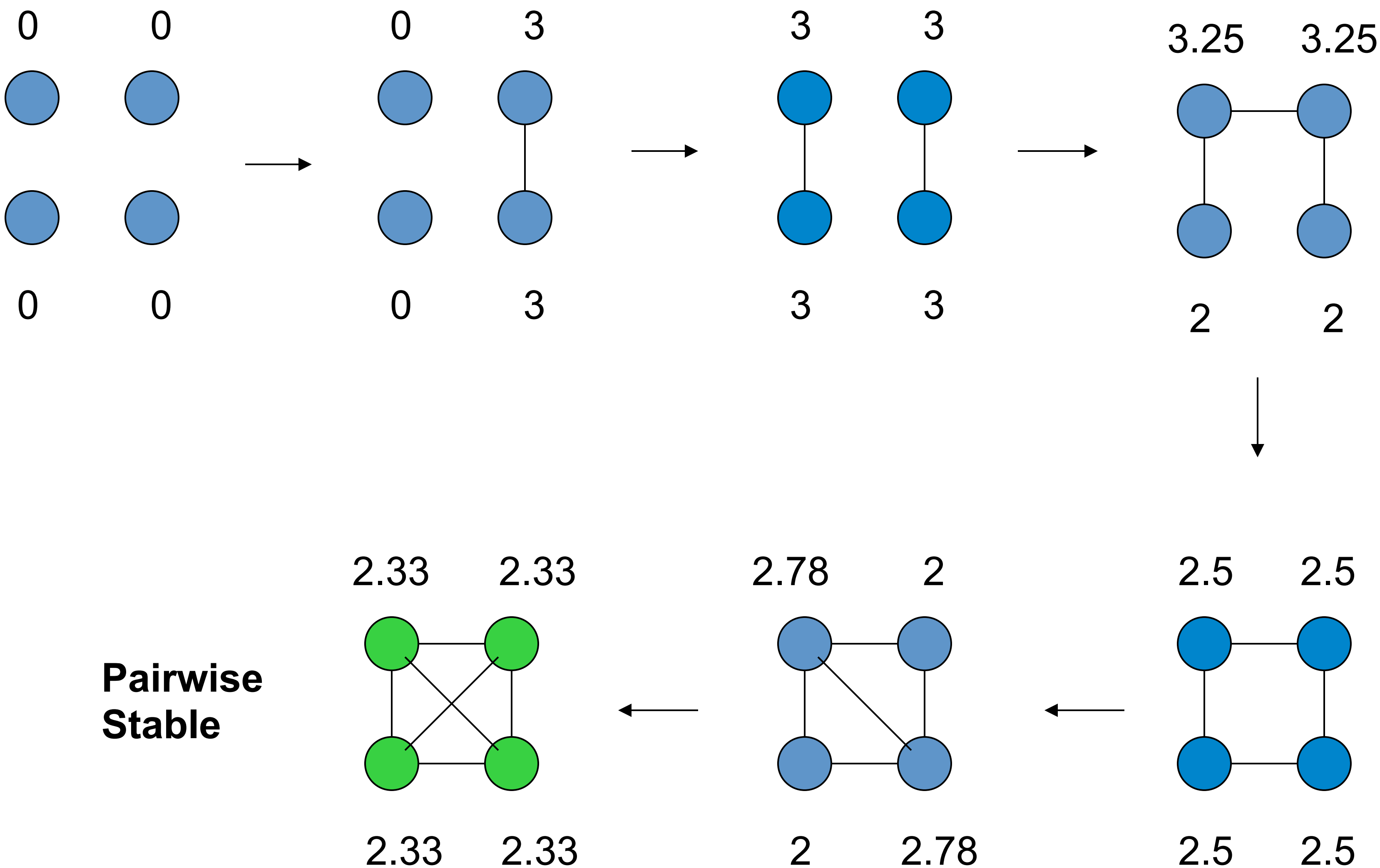
Pairwise Stability



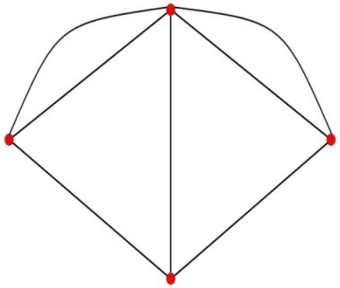
- $u_i(g) \geq u_i(g\text{-}ij)$ for i and ij in g
  - no agent gains from severing a link

- $u_i(g+ij) > u_i(g)$ implies $u_j(g+ij) < u_j(g)$ for ij not in g
  - no two agents both gain from adding a link (at least one strictly)

- a weak concept, but often narrows things down

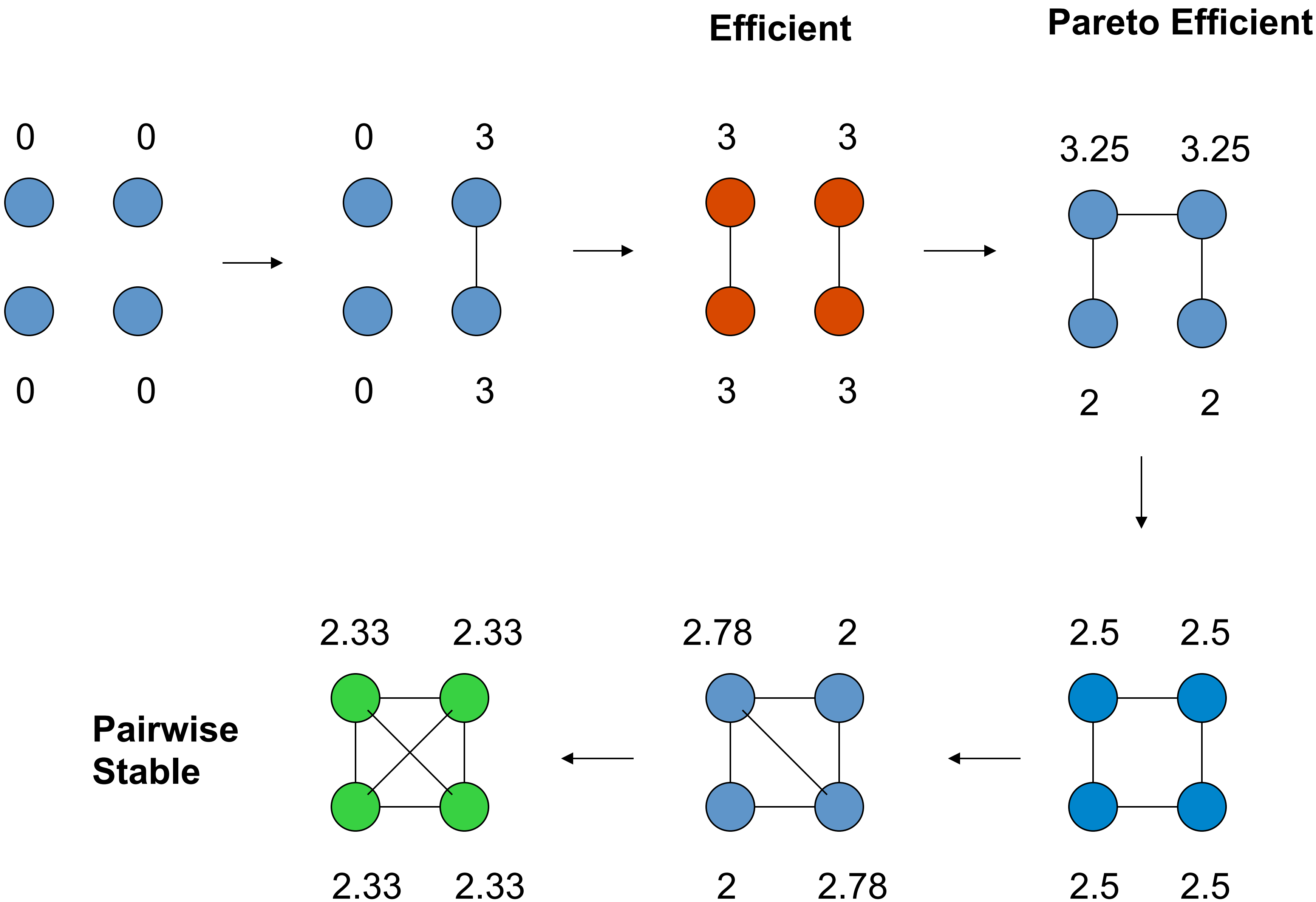

0 0
0 0
0 3
0 3
3 3
3 3
3.25 3.25
2 2
2.5 2.5
2.5 2.5
2.78 2
2 2.78
2.33 2.33
2.33 2.33
Pairwise
Stable

Efficiency

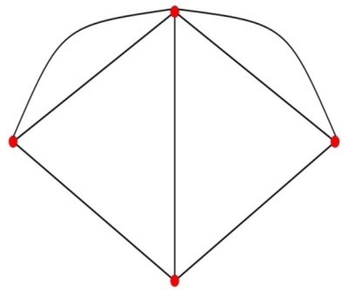


- **Pareto efficient** g: there does not exist g' s.t.
  - $u_i(g') \geq u_i(g)$ for all i, strict for some

- **Efficient** g (Pareto if transfers):
  - g maximizes $\sum u_i(g')$


Efficient
Pareto Efficient
0 0
0 0
0 3
0 3
3 3
3 3
3.25 3.25
2 2
2.5 2.5
2.5 2.5
2.78 2
2 2.78
2.33 2.33
2.33 2.33
Pairwise
Stable

Connections Model JW96



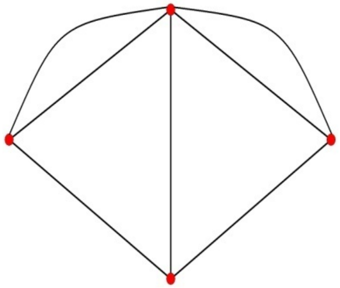
- $0 \leq \delta_{ij} \leq 1$ a benefit parameter for i from path connection between i and j

- $0 \leq c_{ij}$ cost to i of link to j

- $\ell(i,j)$ shortest path length between i,j

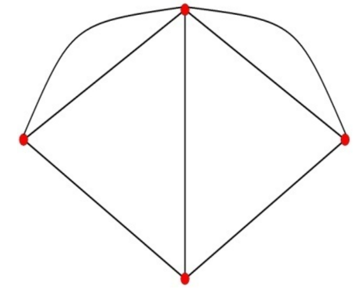
$$u_i(g) = \sum_j \delta_{ij}^{\ \ell(i,j)} - \sum_{j \text{ in } N_i(g)} c_{ij}$$

BJ05 variation: Distance Based Utility Model



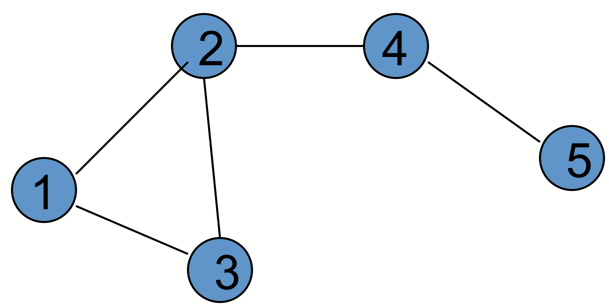
- Let b be a decreasing function
- $u_i(g) = \sum_j b(\ell(i,j)) - d_i(g)\, c$
- $\ell(i,j)$ distance between nodes

Symmetric Version:



- benefit from a friend is b(1)
- benefit from a friend of a friend is b(2)<b(1),...
- cost of a link is c>0

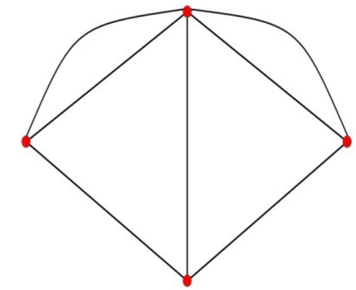
$u_2 = 3b(1) + b(2) - 3c$



$u_1 = 2b(1) + b(2) + b(3) - 2c$

$u_5 = b(1) + b(2) + 2\,b(3) - c$

Efficient Networks in the Symmetric Connections Model

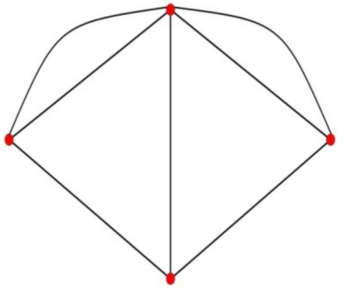


- low cost: $c < b(1)-b(2)$
  - complete network is uniquely efficient

- medium cost: $b(1)-b(2) < c < b(1)+(n-2)b(2)/2$
  - star networks with all agents are uniqely efficient

- high cost: $b(1)+(n-2)b(2)/2 < c$
  - empty network is uniquely efficient

Proof

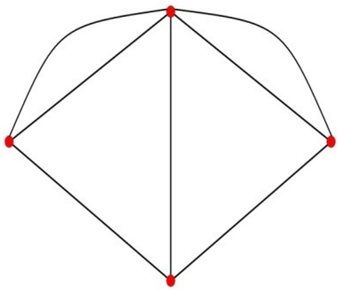


- $c < b(1)-b(2)$ then $u_i(g+ij) > u_i(g)$ if $ij \in g$

Also $u_k(g+ij) \geq u_k(g)$ if $ij \in g$ for every k, thus

$\Sigma_k u_k(g+ij) > \Sigma_k u_k(g)$

- $c > b(1)-b(2)$ first, show that the value of a component is highest when the component is a star



- value of a star with k players is

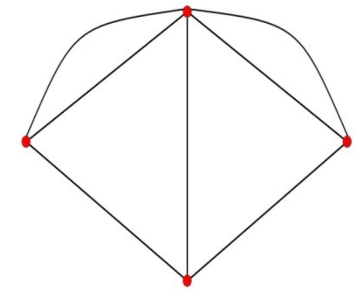
  $2(k-1)\ [\ b(1) - c] + (k-1)(k-2)b(2)$

- value of a network with k players and m links ($m \geq k-1$) is at most

  $2m\ [b(1) - c]\ + [k(k-1)-2m]b(2)$

- difference is

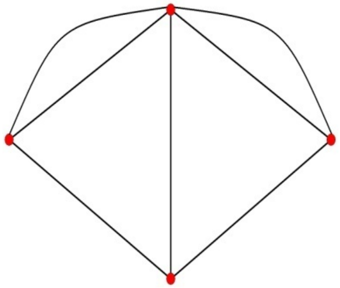
  $2(m-(k-1))\ [b(2) -(b(1) - c)]\ > 0$ if $m > k-1$



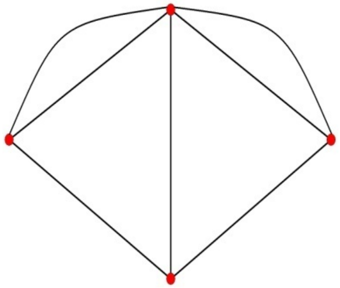
- If m = k-1 and not a star, then some pair is at a distance of more than 2, so less value than a star:
- value of a star with k players is

  2(k-1) [b(1) – c] + (k-1)(k-2)b(2)
- value of a component with k players and k-1 links that is not a star is at most

  2(k-1) [b(1) – c] +[(k-1)(k-2)-1]b(2) +b(3)
- Star is better



- Check that if two separate star components each generate nonnegative utility, then one star with all those players generates higher utility
- Separate: 2(k-1) [b(1) – c] + (k-1)(k-2)b(2) + 2(k'-1) [b(1) – c] + (k'-1)(k'-2)b(2)
- = 2(k+k'-2) [b(1) – c] + [(k-1)(k-2)+ (k'-1)(k'-2)]b(2)
- As one star: 2(k+k'-1) [b(1) – c] + (k+k'-1)(k+k'-2)b(2)
- second expression is greater...

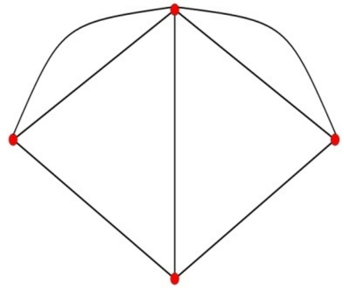


- So efficient networks are collections of stars or empty networks
- So, either a star with all players or empty:

- Want a star if its value is >0, so when

$$2(n-1)\,[\,b(1) - c] + (n-1)(n-2)b(2) > 0$$

Pairwise Stability



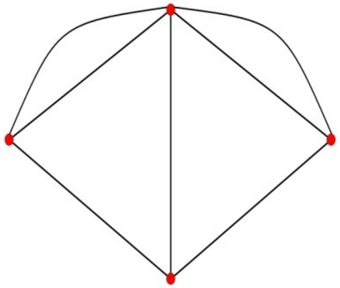
- low cost: $c < b(1)-b(2)$
  - complete network is pairwise stable

- medium/low cost: $b(1)-b(2) < c < b(1)$
  - star network is pairwise stable
  - others are also pairwise stable

- medium/high cost: $b(1) < c < b(1)+(n-2)b(2)/2$
  - star network is not pairwise stable (no loose ends)
  - nonempty pairwise stable networks are over-connected and may include too few agents

- high cost: $b(1)+(n-2)b(2)/2 < c$
  - empty network is pairwise stable

Inefficiency:

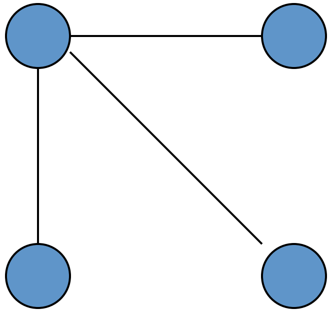


- payoff to center:

  $3b(1) - 3c$

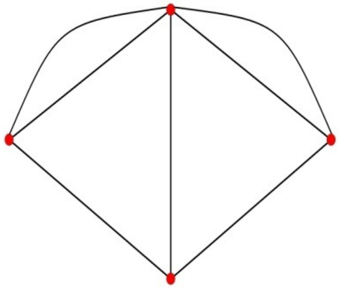
  not pairwise stable if

  $b(1) < c$



- only sustain link if it brings indirect benefits

Inefficiency:

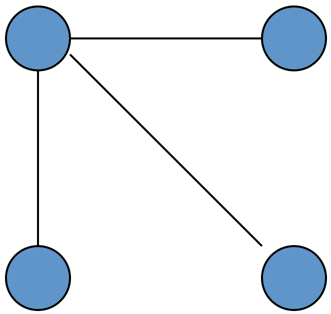


- payoff to center:

  $3b(1) - 3c$

  not pairwise stable if

  $b(1) < c$

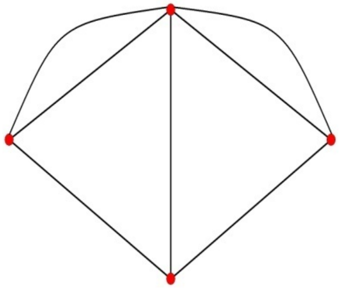


Overall payoff: $6b(1) + 6b(2) - 6c$

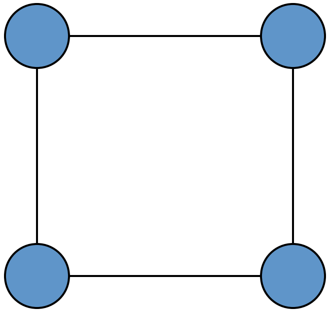
Peripheral players gain indirect benefits

Center player does not account for them

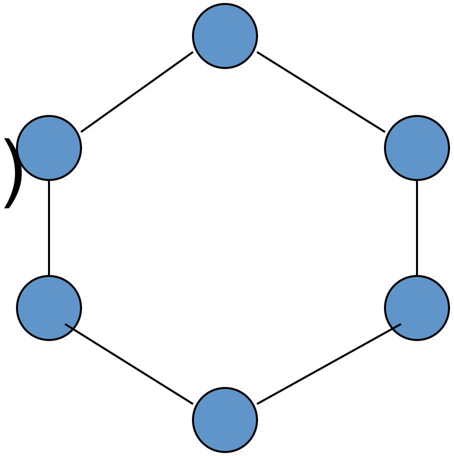
Example: Pairwise stable and inefficient



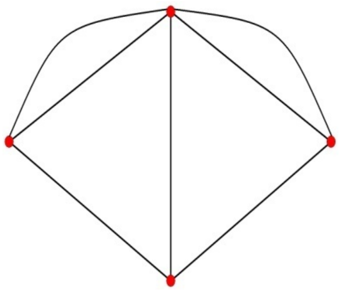
- b(1)-b(2) < c < b(1)-b(3), n=4



- b(1)-b(3)< c < (**b(1)+b(2)+b(3)**)(1- b(2))

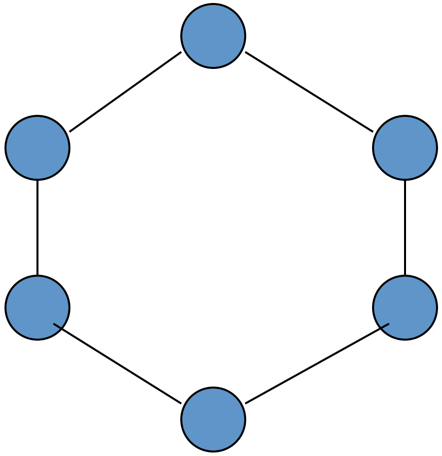


Example: Pairwise stable and inefficient

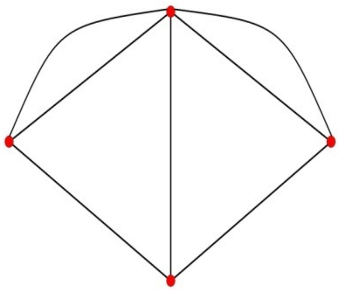


unique nonempty pairwise stable network architecture if

$b(1) < c < (b(1)+b(2)+b(3))(1- b(2))$, n=6



Externalities



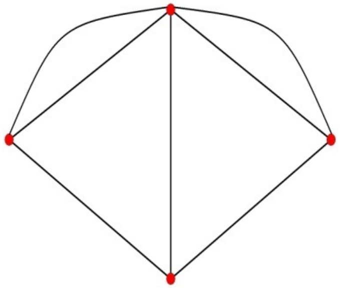
- Positive:

  $u_k(g+ij) \geq u_k(g)$ if $ij \in g$ for every $k \neq i,j$

- Negative:

  $u_k(g+ij) \leq u_k(g)$ if $ij \in g$ for every $k \neq i,j$

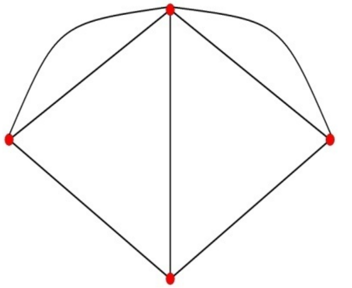
Externalities



- Inefficiency in connections model due to positive externalities - ``no loose ends''

- What about models with negative externalities?

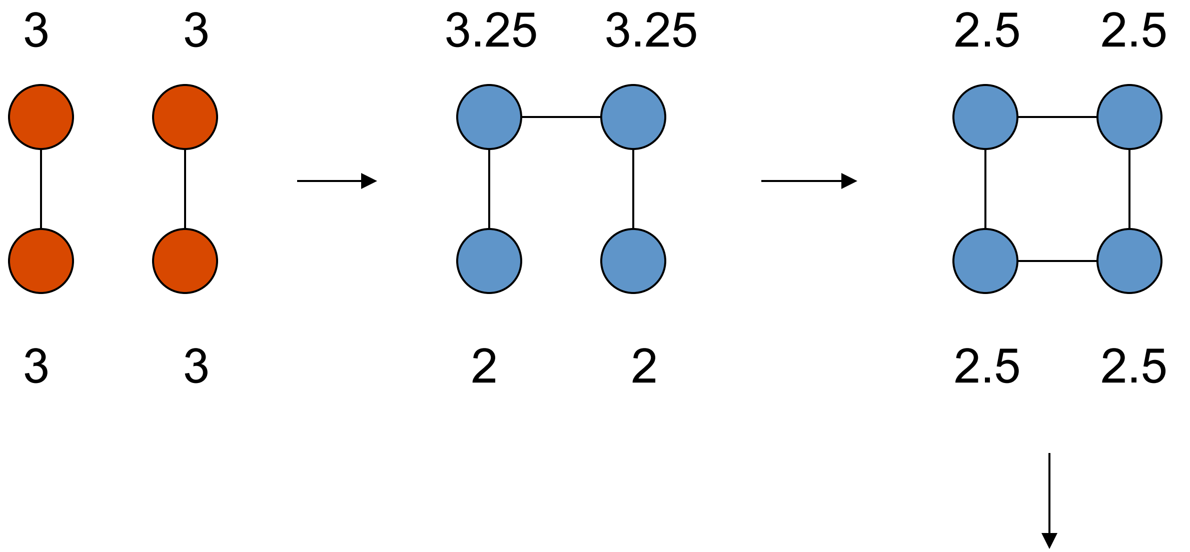
Example: ``Coauthor'' JW96



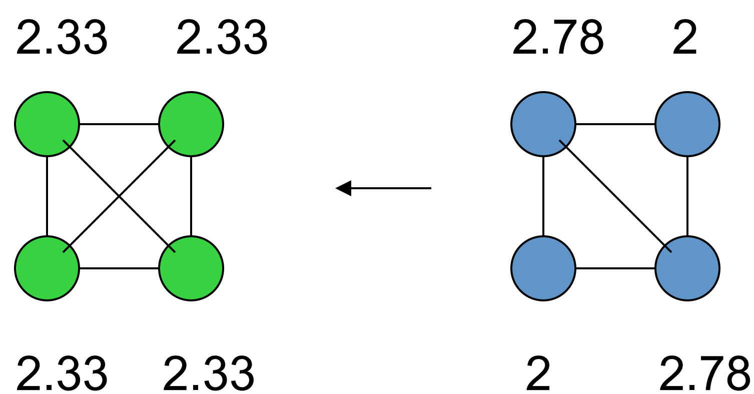
- Agents get value from research collaboration
  - value for each relationship depends on time each puts into it
  - plus an interaction term, which is product of the times spent

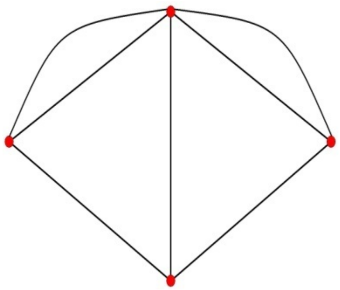
$$u_i(g) = \sum_{j:\, ij \in g} [1/d_i + 1/d_j + 1/(d_i d_j)]$$
$$= 1 + \sum_{j:\, ij \in g} [1/d_j + 1/(d_i d_j)]$$

Efficient:



Pairwise Stable:

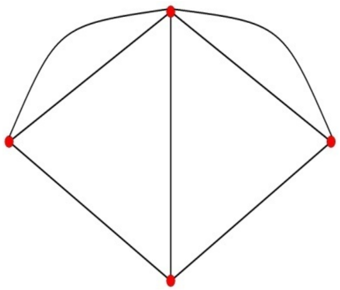




- no direct costs to link

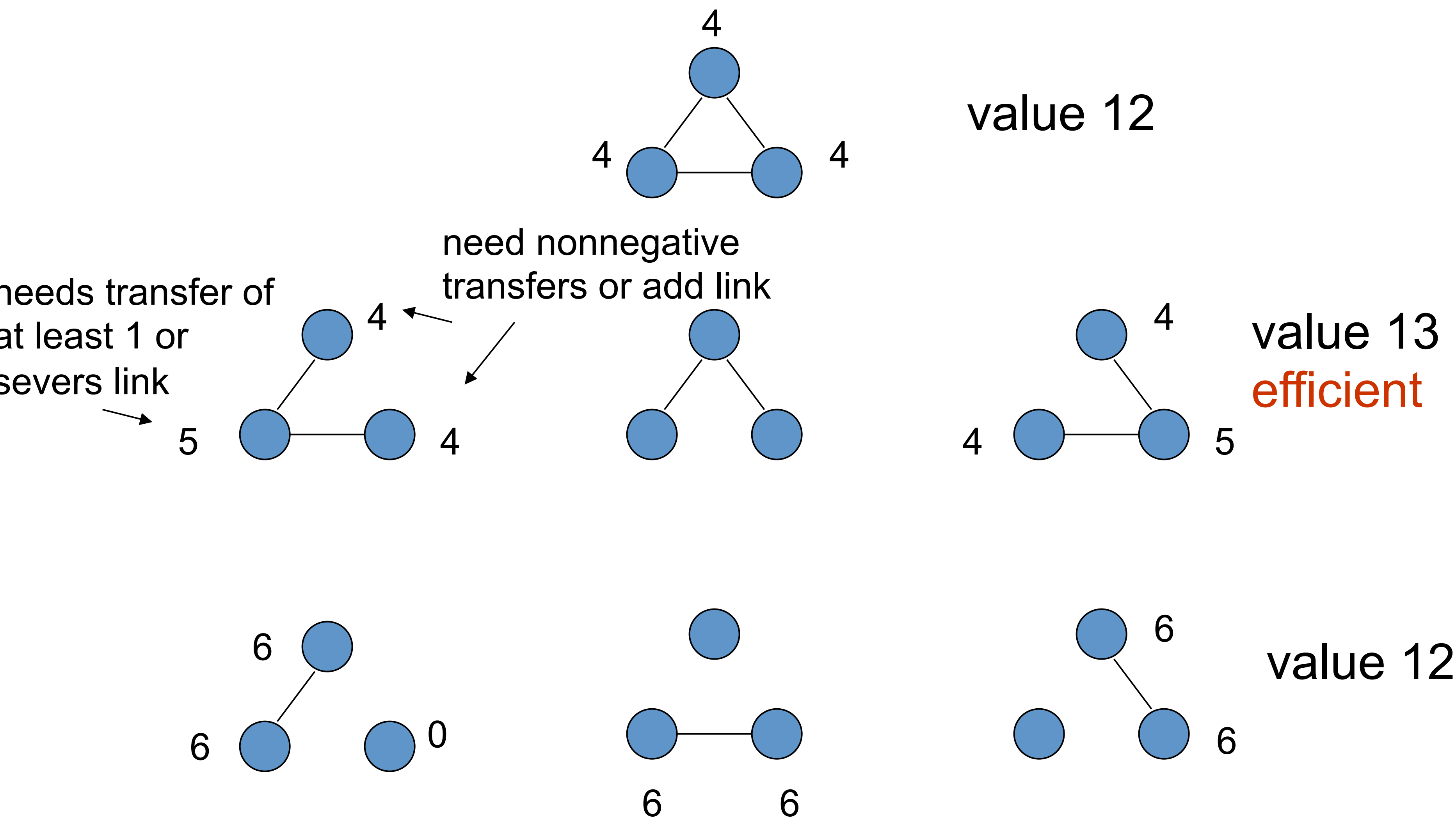
- n is even:
  - efficient networks: pairs
  - pairwise stable networks consist of completely connected components, each of a different size, one has more than the square of the number of nodes in the other
  - by adding a link, dilute existing synergies, only add if new coauthor brings comparable worth

Stable and Efficient only coincide in special cases

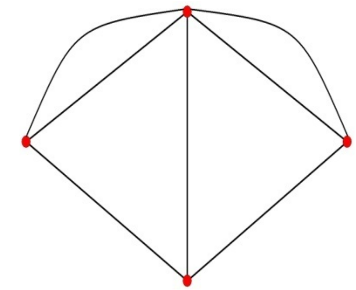


- Can transfers help in other cases?
- What can we say about when conflict exists?
- What can we say about when transfers improve efficiency?
- Are transfers in players’ interests?

Transfers cannot always help (JW 1996)



Enriching Such Models



- Costs depend on geography and characteristics of nodes
  - easier to be friends with neighbors
  - easier to relate to people with similar background
- Benefits depend on characteristics of nodes
  - synergies from working together, trading, sharing risk, exchanging favors..
  - complementarities: benefits from diversity...
- Some randomness in who meets whom
  - models that combine cost/benefit/choice with chance

Can economic models match observables?

- Small worlds derived from costs/benefits
  - low costs to local links – high clustering
  - high value to distant connections – low diameter
  - high cost of distant connections – few distant links

Geographic Connections (Johnson-Gilles (2000), Carayol-Roux (2005), Jackson-Rogers (2005), Galeotti-Goyal- Kamphorst (2006),...)

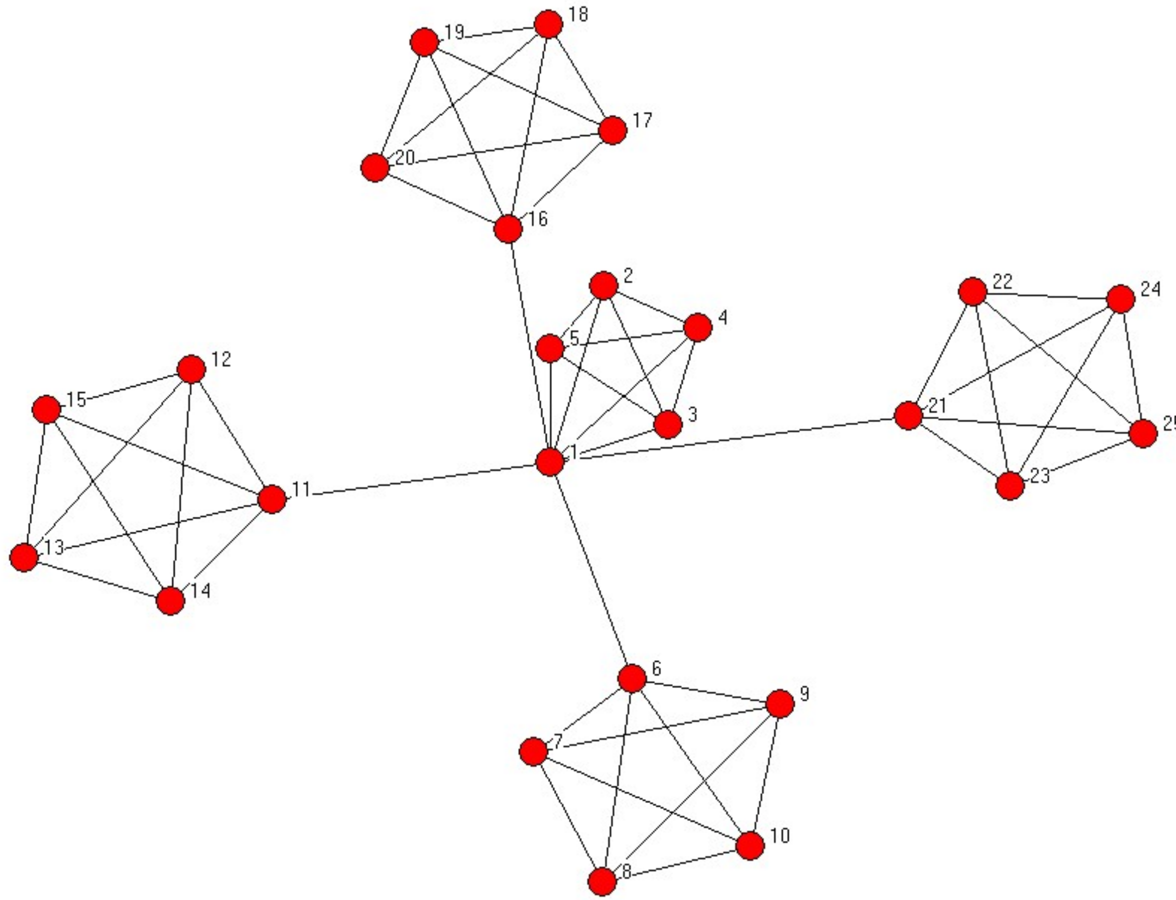
Islands connections model JR05

- J players live on an island, K islands
- cost c of link to player on the island
- cost C>c of link to player on another island

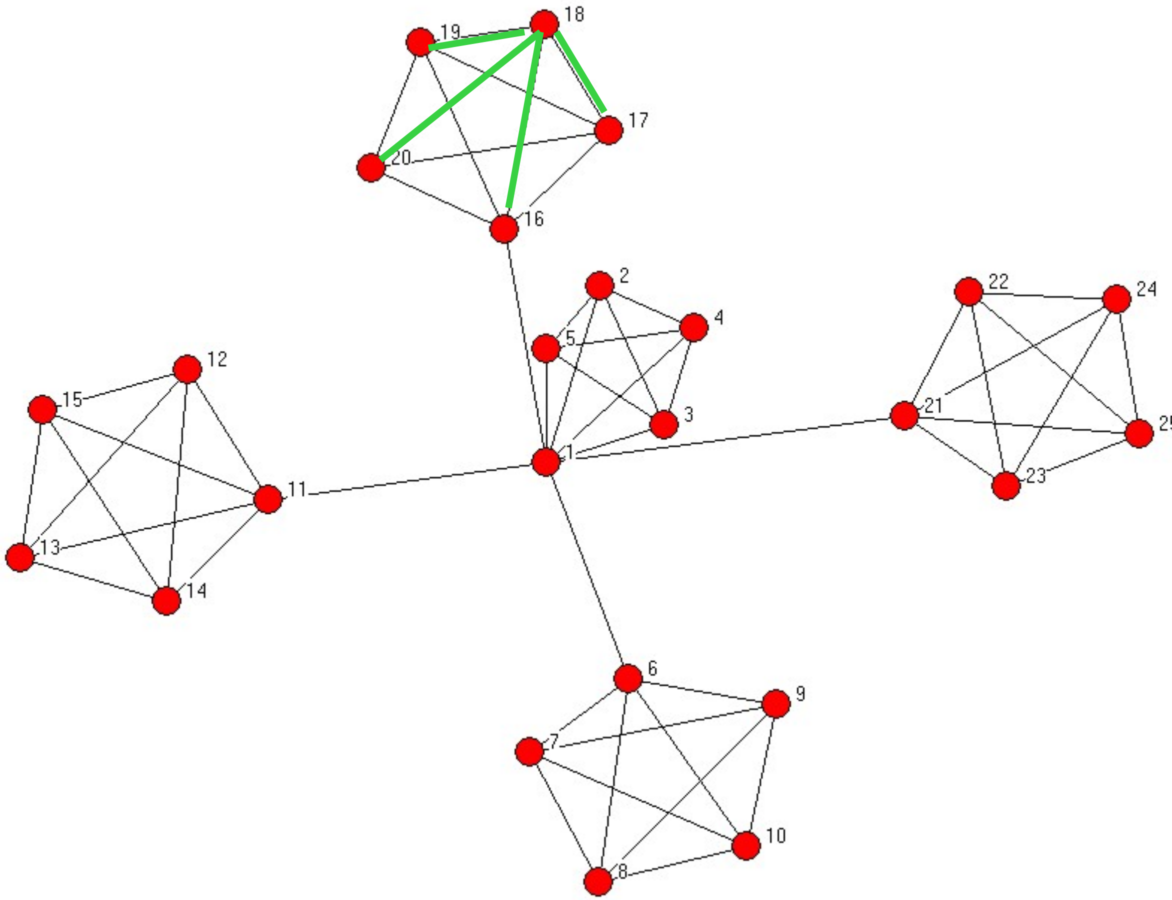
Results:

- High clustering within islands, few links across
- small distances

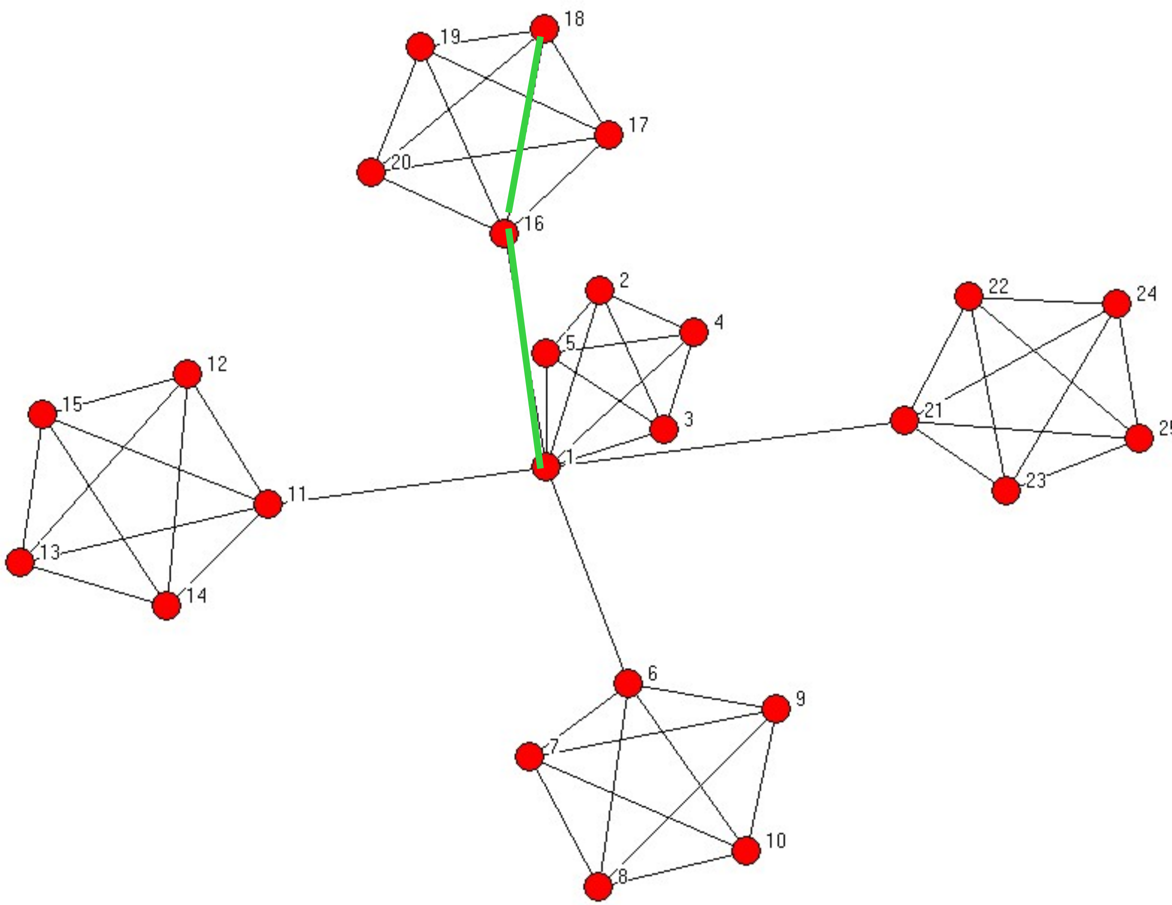
$$4\delta - 4c + \delta^2 + 7\delta^3 + 12\,\delta^4$$



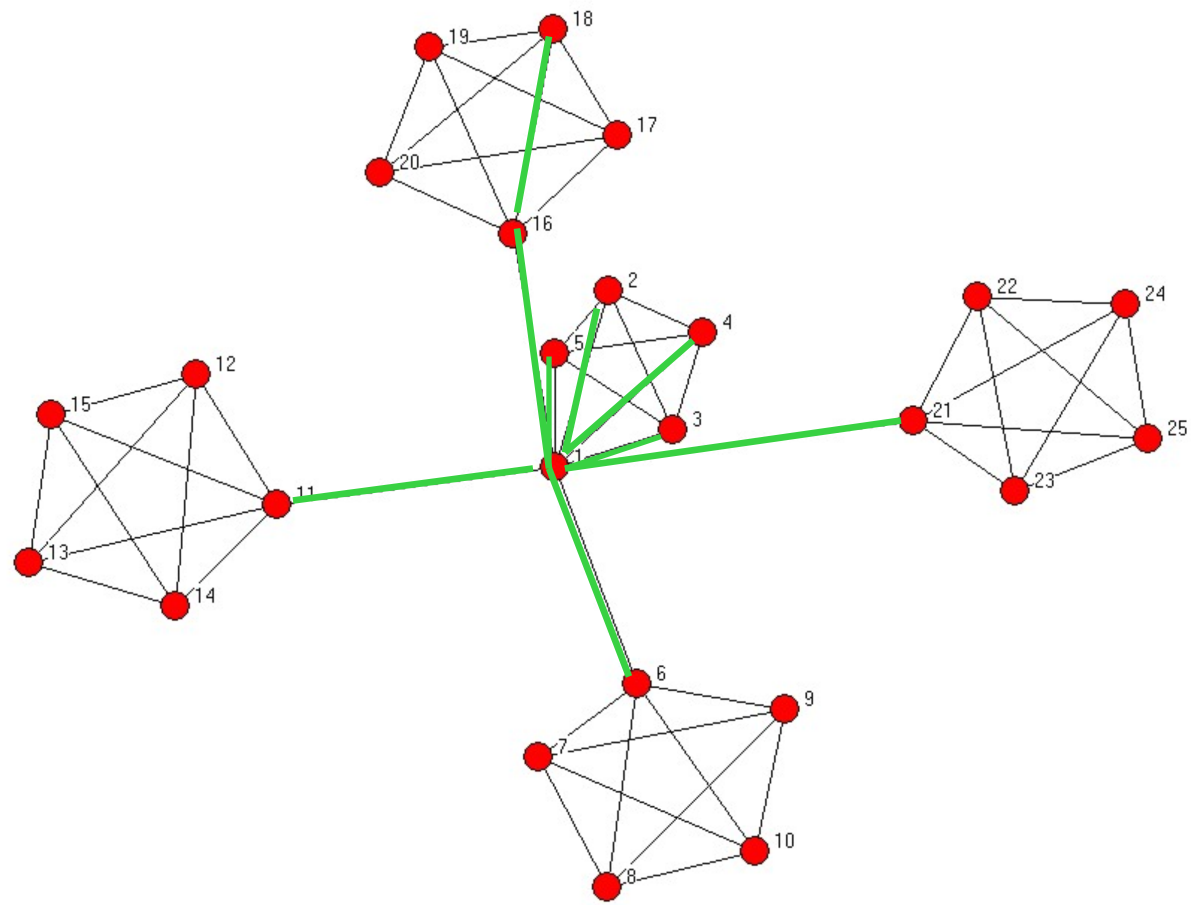
$4\delta - 4c + \delta^2 + 7\delta^3 + 12\,\delta^4$



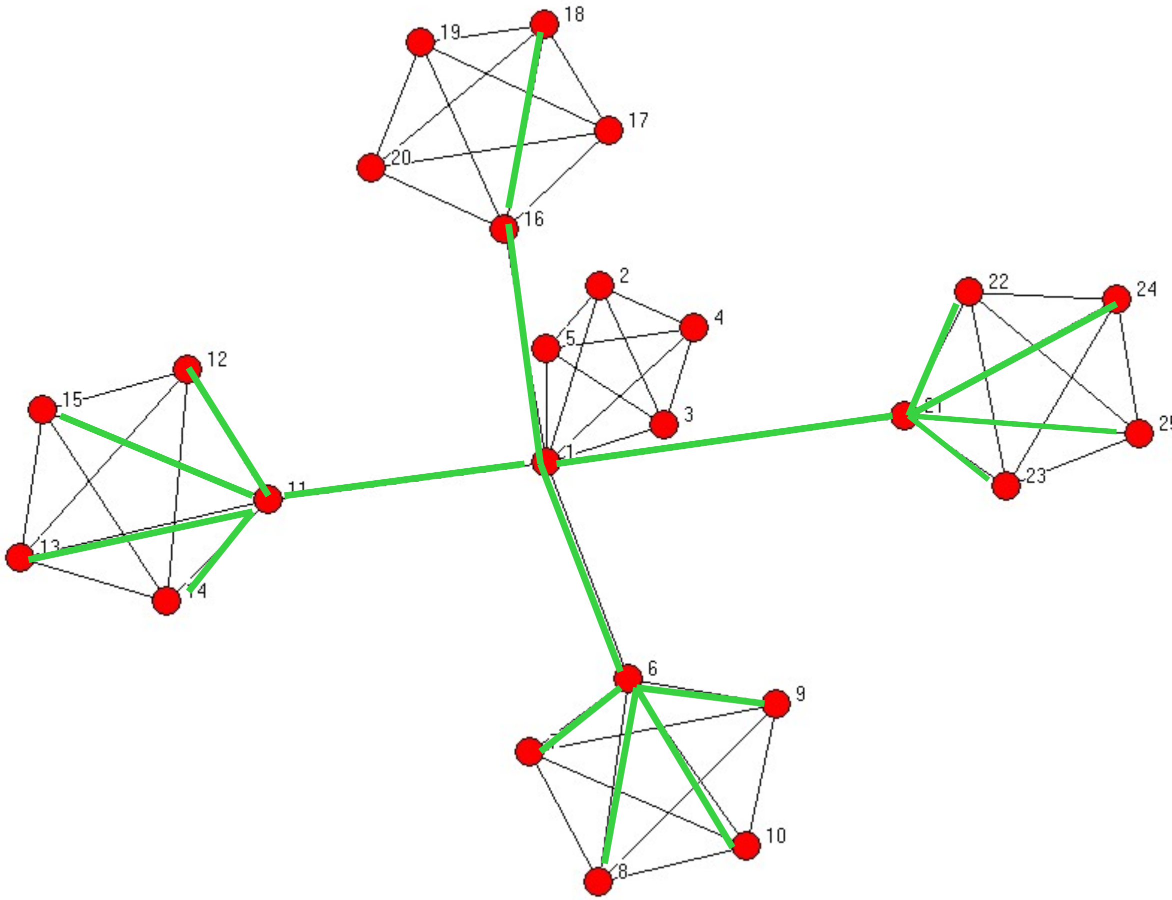
$$4\delta - 4c + \delta^2 + 7\delta^3 + 12\,\delta^4$$



$4\delta \; - 4c + \delta^2 + 7\delta^3 + 12\,\delta^4$

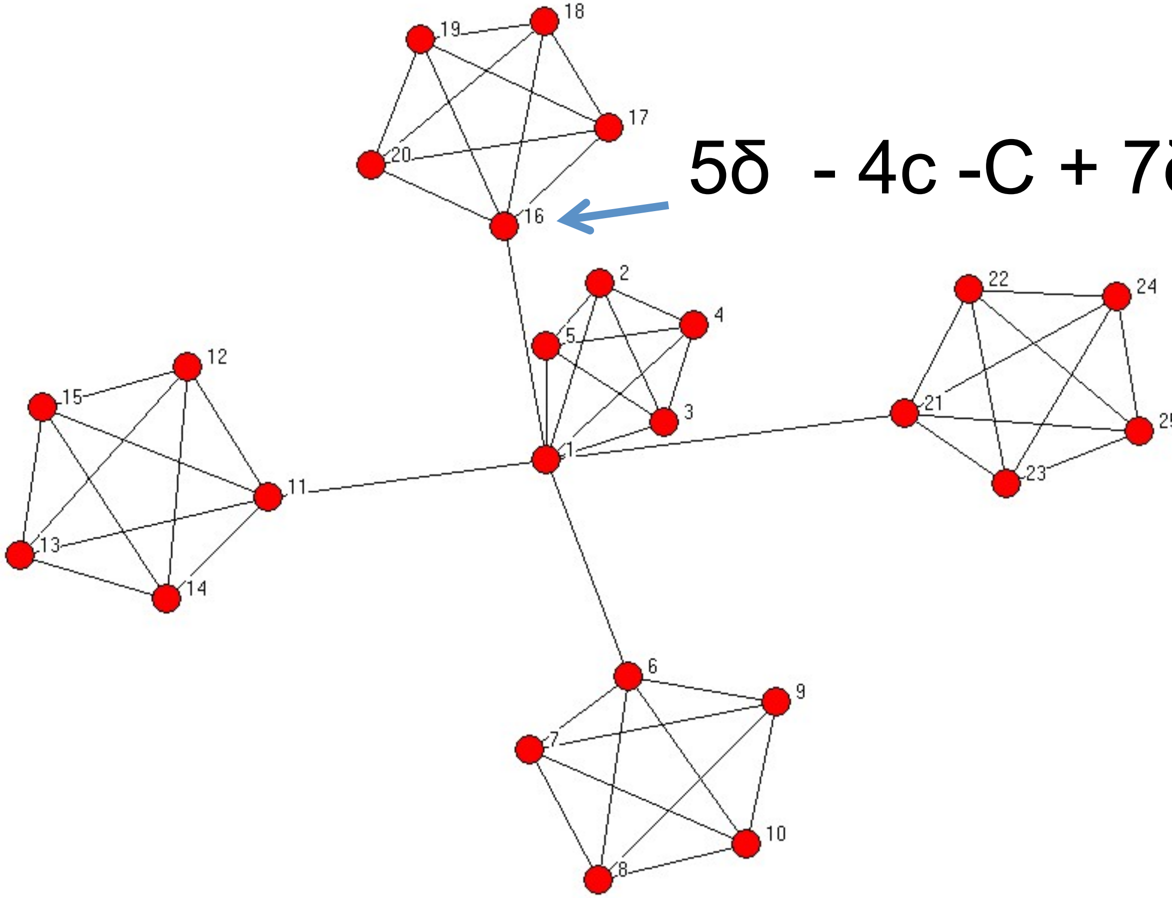


$$4\delta - 4c + \delta^2 + 7\delta^3 + 12\,\delta^4$$



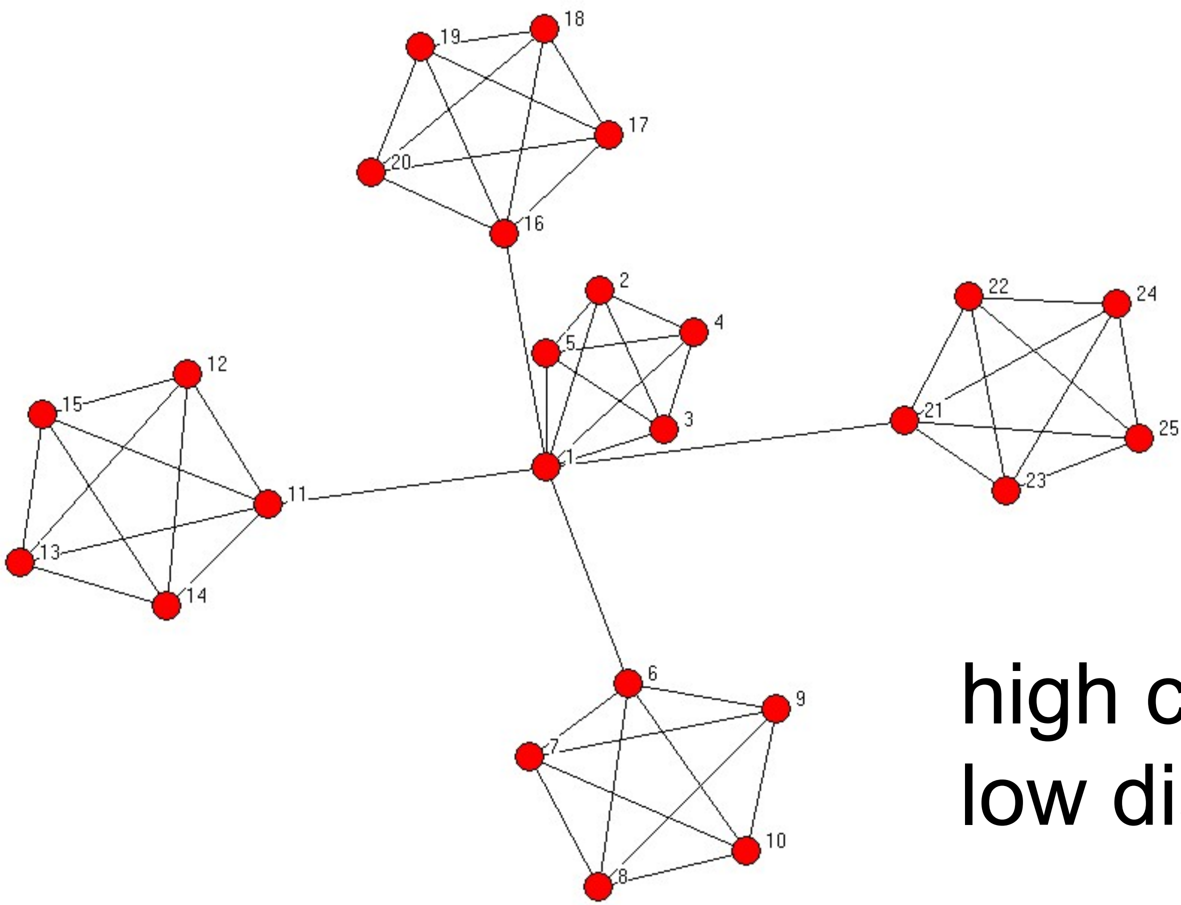
$4\delta \ - 4c + \delta^2 + 7\delta^3 + 12\ \delta^4$

$5\delta \ - 4c - C + 7\delta^2 + 12\delta^3$



low cost of link to player on own ``island''
– high cost across islands

Pairwise stable: (c<.04, 1<C<4.5, δ=.95)

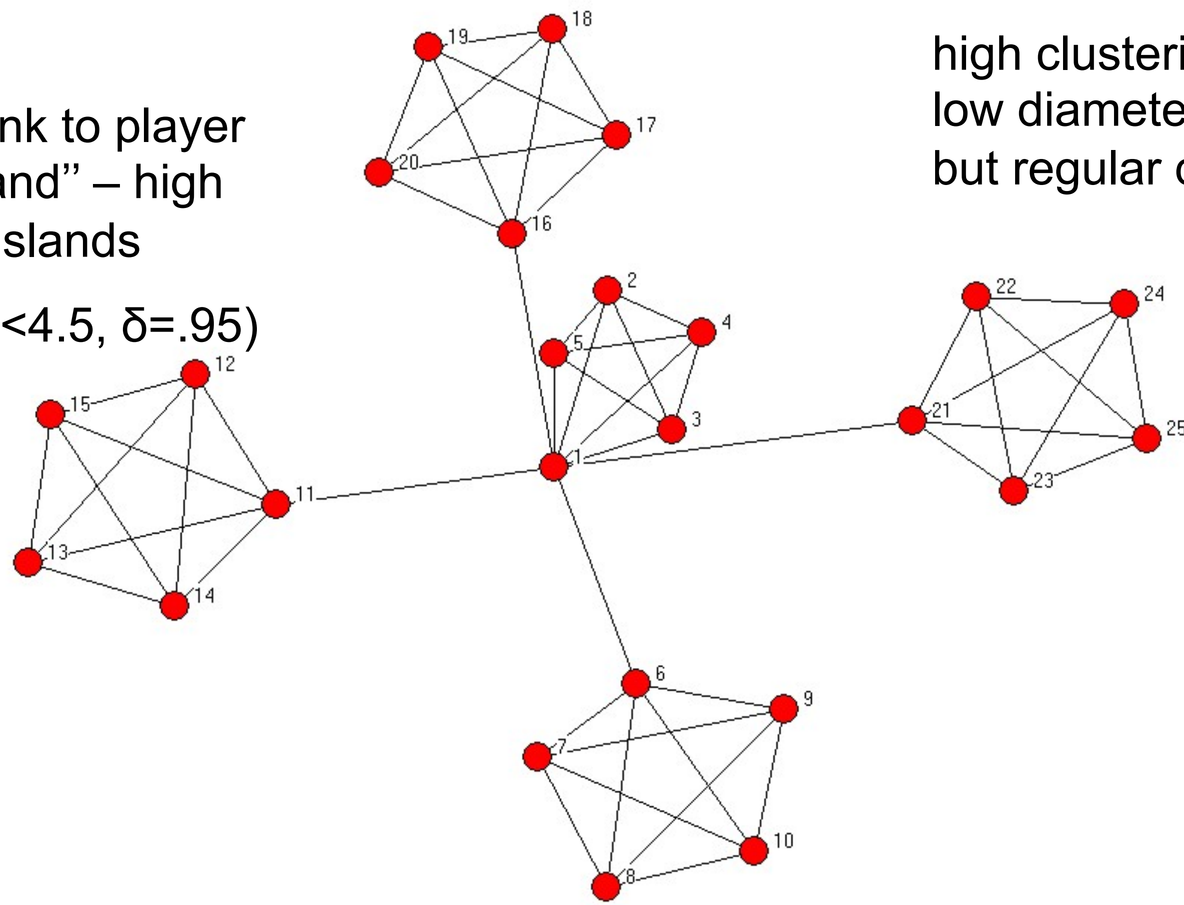


high clustering,
low diameter,

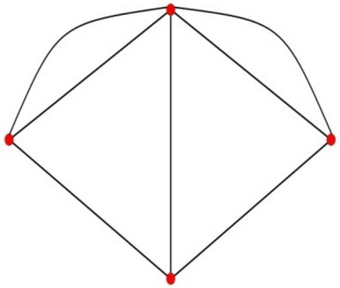
low cost of link to player on own ``island'' – high cost across islands

(c<.04, 1<C<4.5, δ=.95)

high clustering,
low diameter,
but regular degree



Carayol-Roux (2007)



- Simple geographic version of the connections model

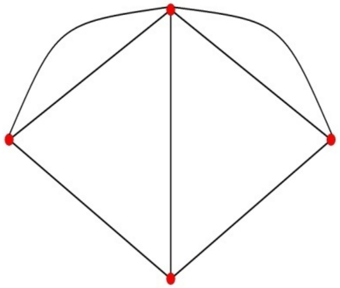
- 20 nodes located on a circle

- cost of linking from node i to j is

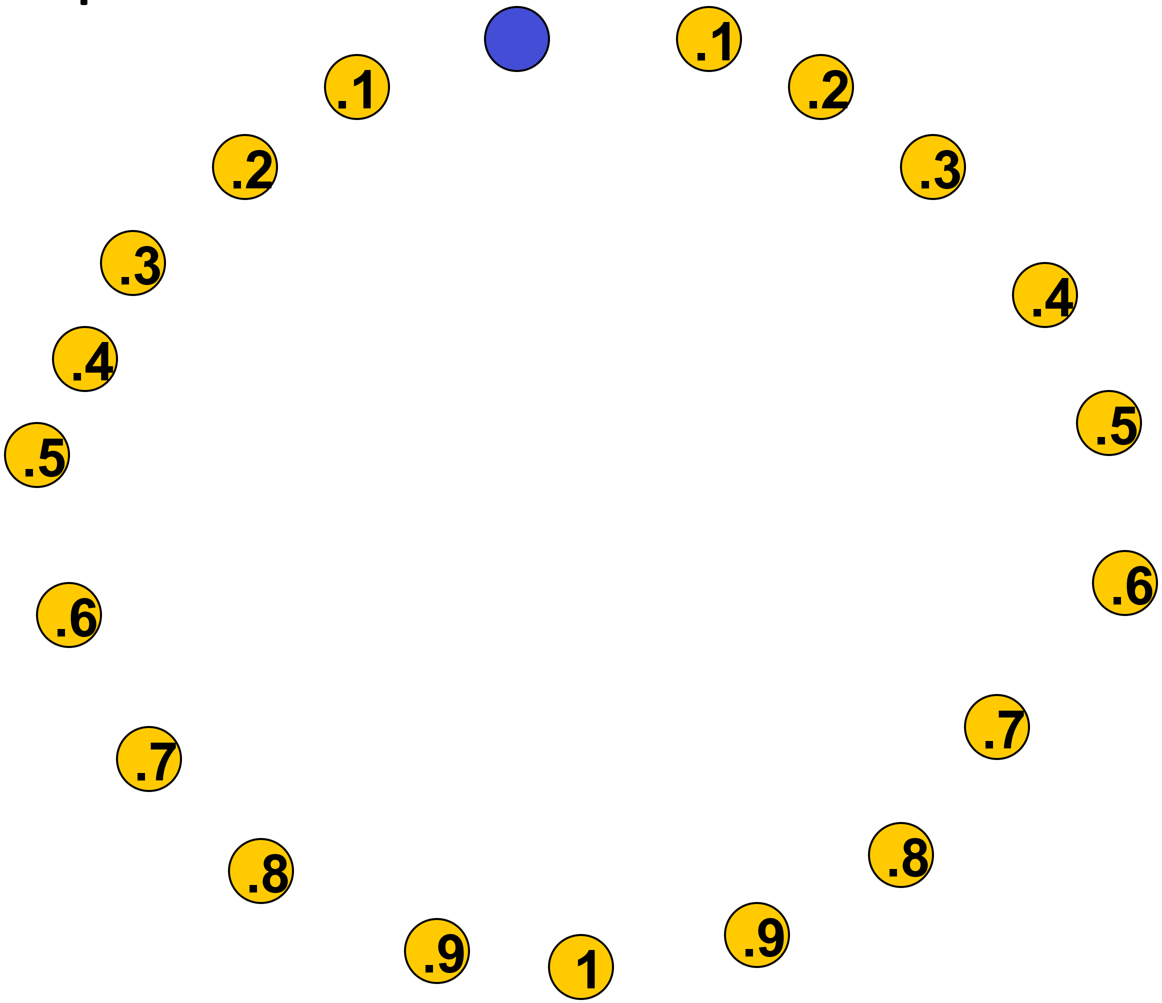
  dist on circle / 10

- e.g., cost is 7/10 to connect nodes 2 and 9

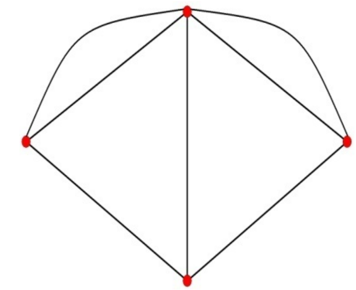
Another Geographic Cost:



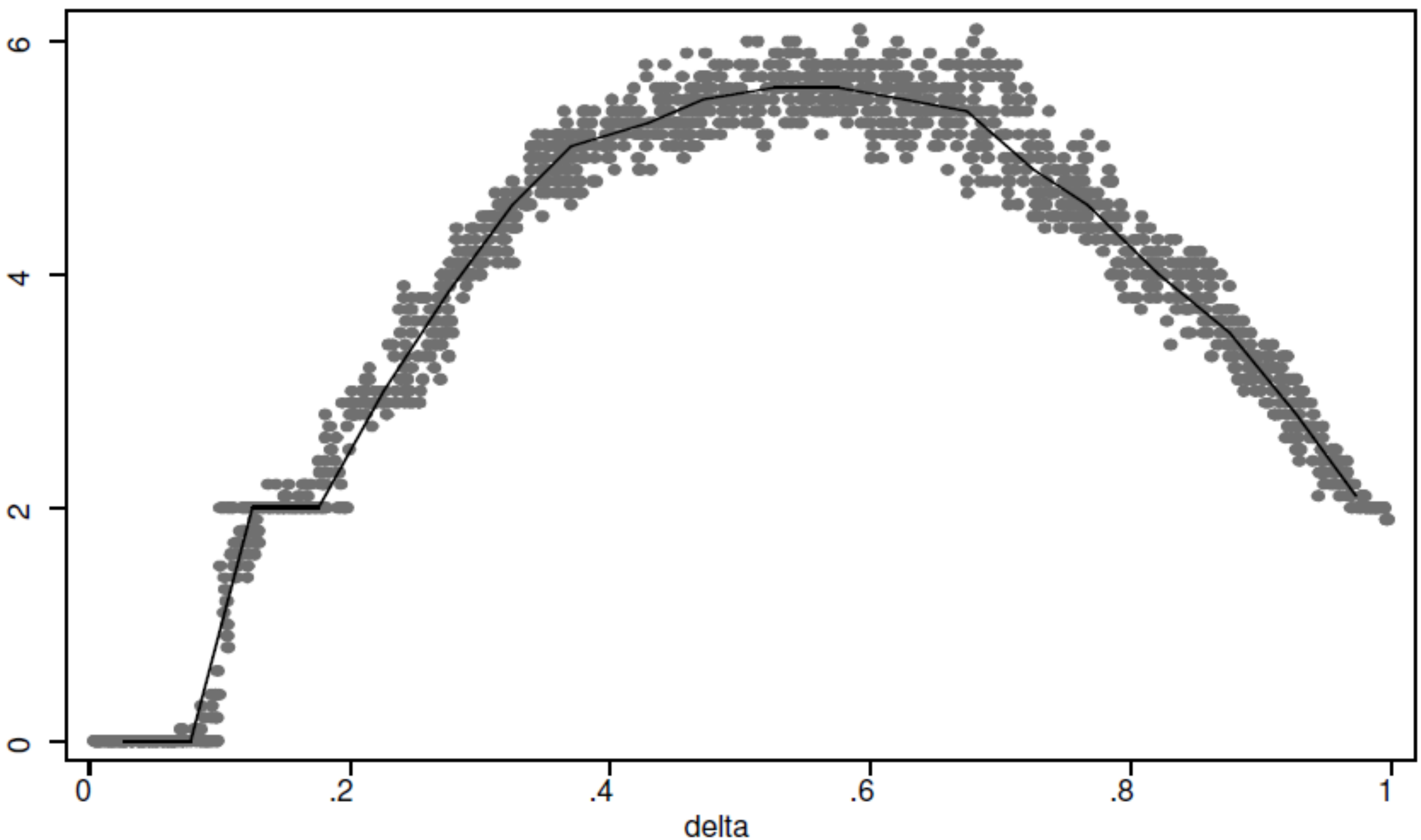
- cost proportional to distance

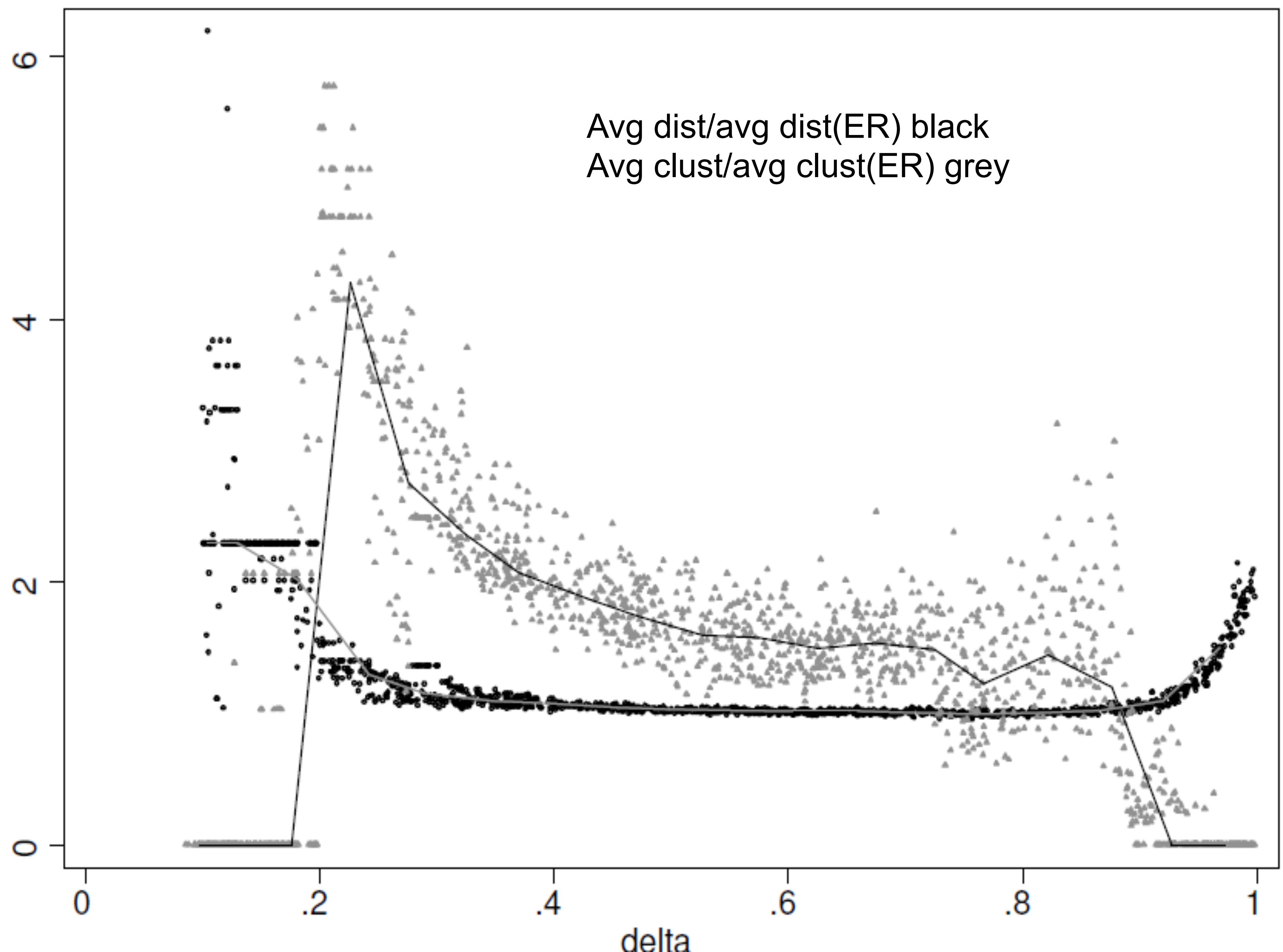


Simulations Carayol-Roux (2007) average degree:



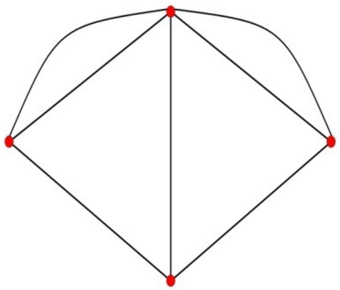
20 Agents on a circle, cost = distance/(n/2), average over 1500 simulations random order of links to form improving paths + noise





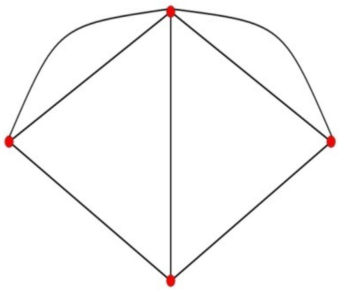
Avg dist/avg dist(ER) black
Avg clust/avg clust(ER) grey
6
4
2
0
0
.2
.4
.6
.8
1
delta

Approaches



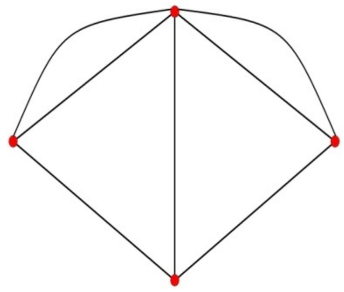
- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Growing Random Networks



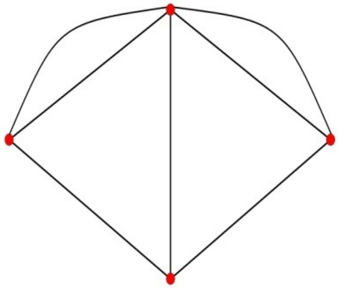
- Citation networks
- Web
- Scientific networks
- Societies...

What do they add?



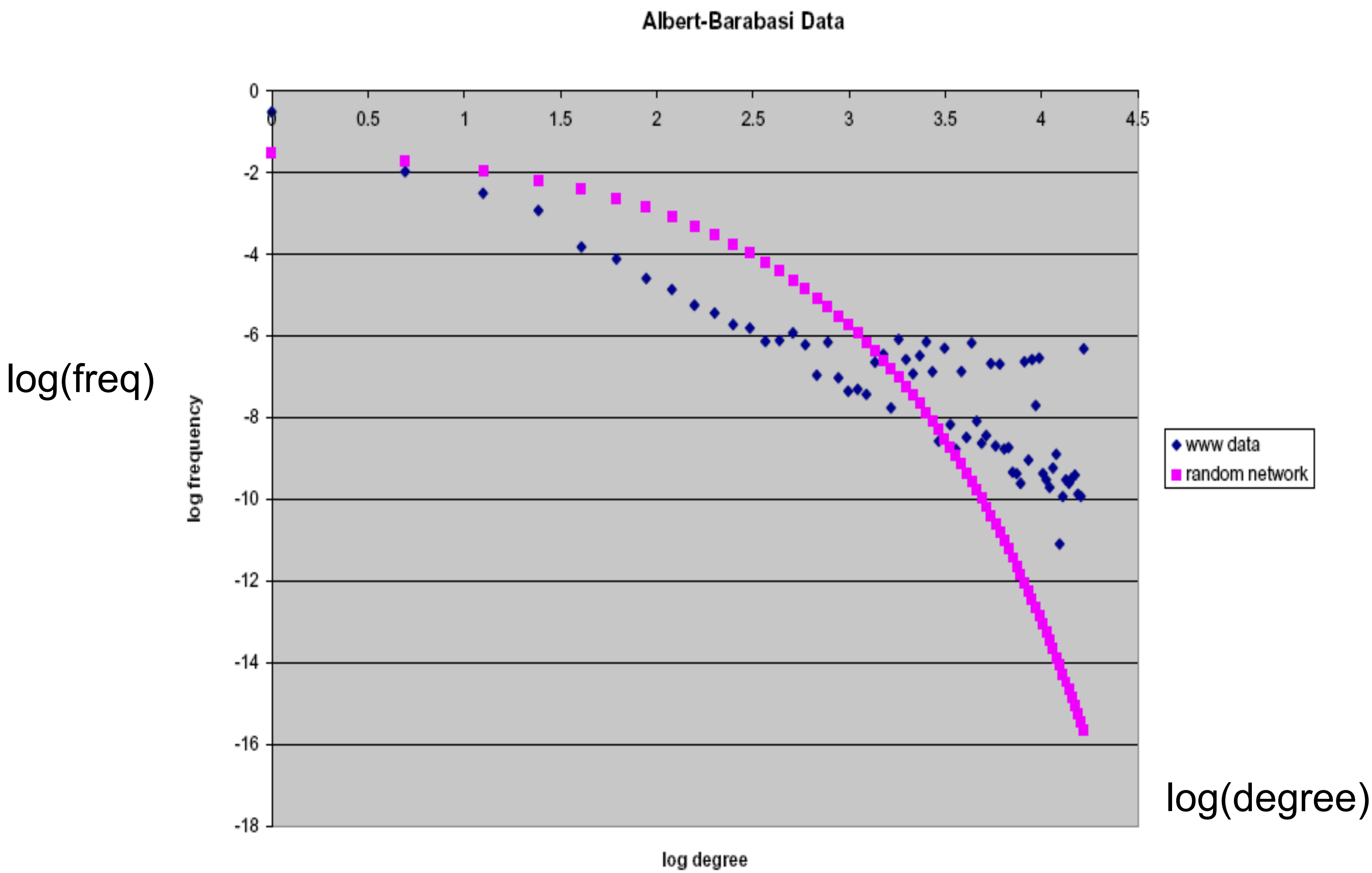
- Realism(?)
- Natural form of heterogeneity via age
- A form of dynamics
- Natural way of varying degree distributions
  - not pre-specified as in static models

Preferential Attachment

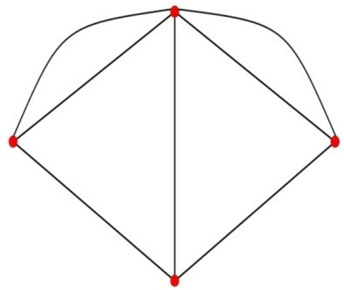


- Other methods of linking
- Can we get other degree distributions: ``Power laws’’?

Degree – ND www Albert, Jeong, Barabasi (1999)



Power Law Explanations

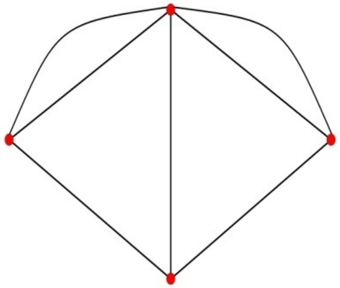


- Simon (1955):

- Rich get richer – growth of existing objects is proportional to size

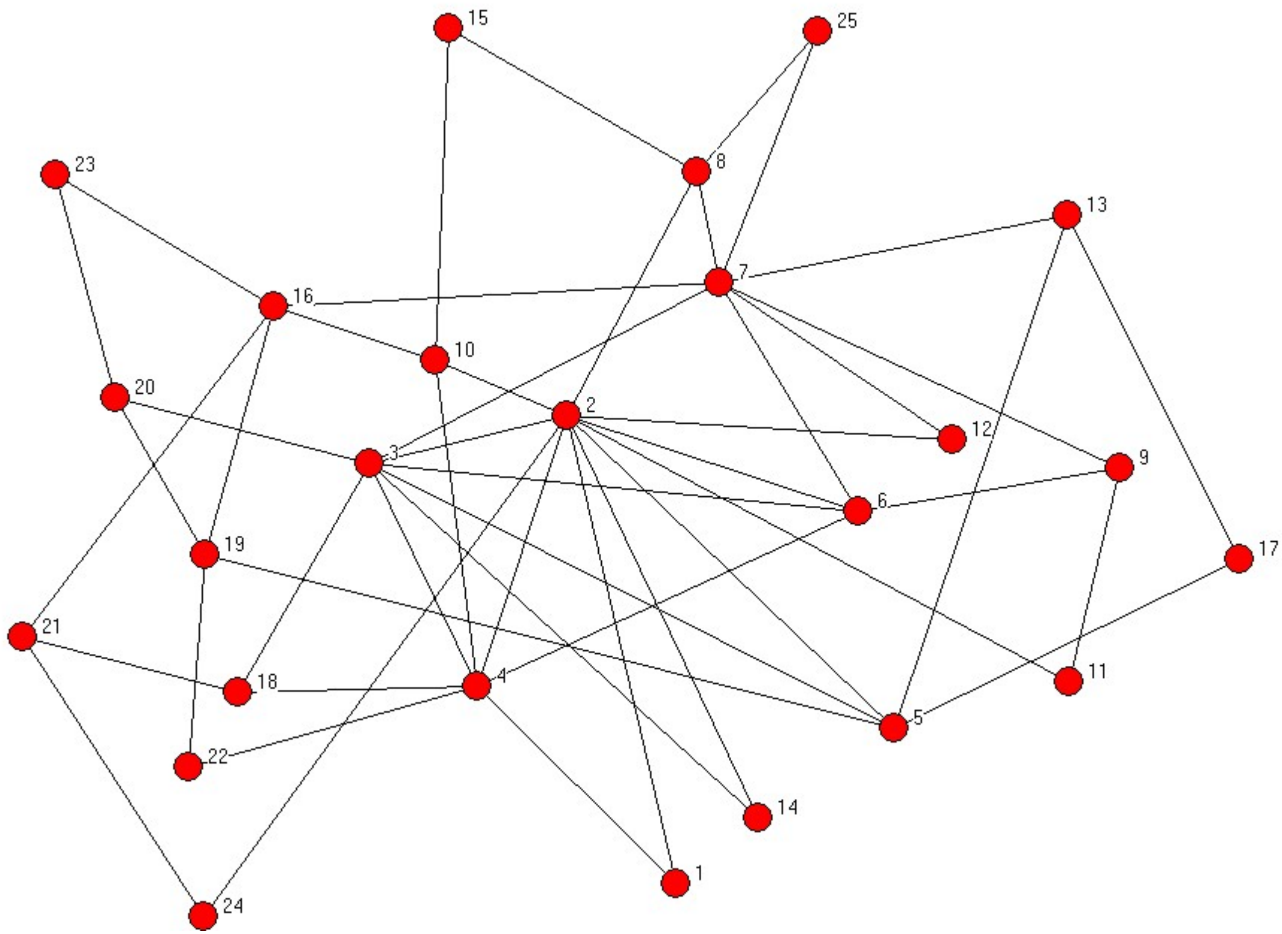
- New objects enter over time

Preferential Attachment (Price (1976), Barabasi and Albert (2001))

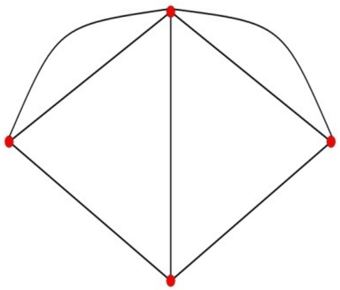


- Previous models don't have the ``fat tails'' of degree distributions
- Nodes born over time, form links at random with existing nodes
  - Form links with probability *proportional to number of links a node already has* - ``rich get richer''

Preferential Attachment (Price (1976), Barabasi and Albert (2001))

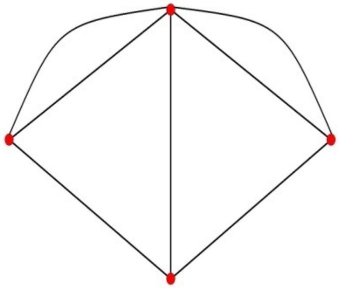


Preferential Attachment



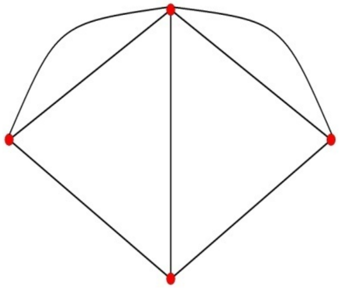
- Newborn nodes form m links to existing nodes
- tm links in total
- total degree is 2tm
- ***Probability of attaching to i is*** $d_i(t)/2tm$

Mean Field Approximation



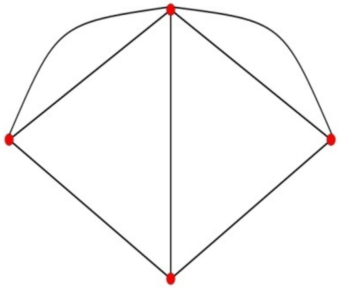
- Continuous time approximation
- Distribution of expected degrees
- Check by simulation??

Distribution of Expected Degrees



- $dd_i(t)/dt = md_i(t)/2tm = d_i(t)/2t$ and $d_i(i)=m$

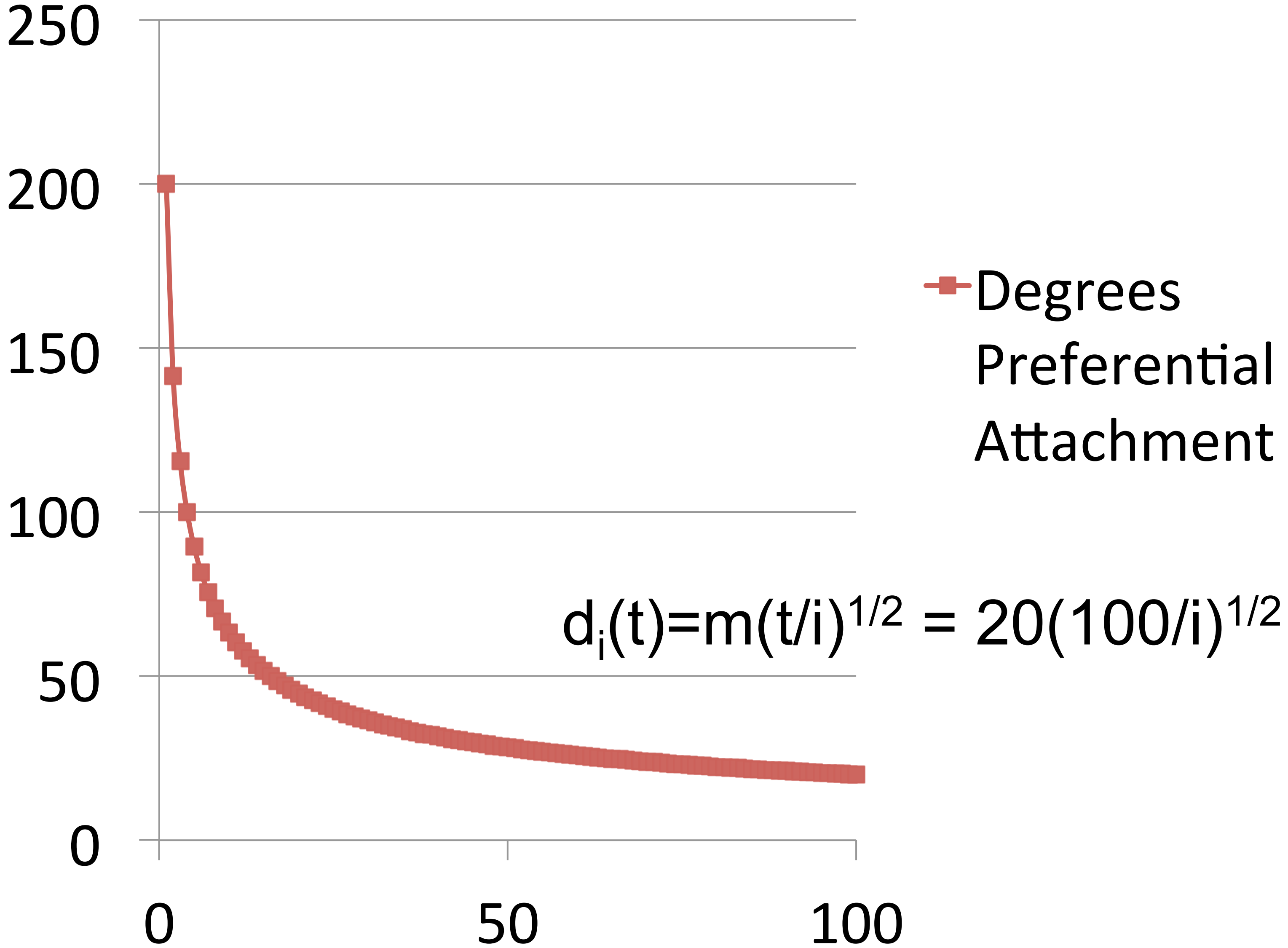
Distribution of Expected Degrees



- $dd_i(t)/dt = md_i(t)/2tm = d_i(t)/2t$ and $d_i(i)=m$

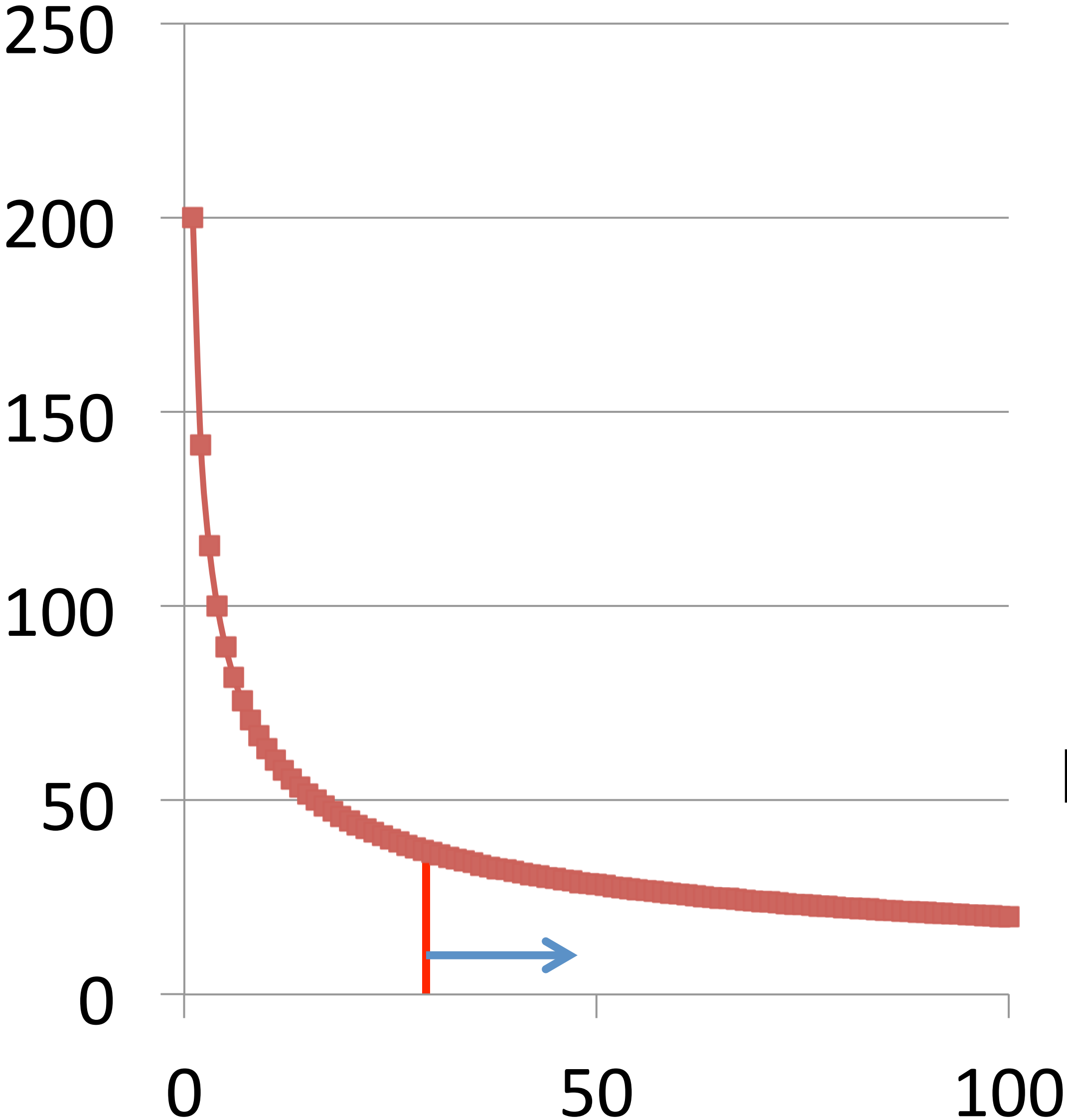
- $d_i(t) = m\,(t/i)^{1/2}$

Degrees Preferential Attachment



$$d_i(t)=m(t/i)^{1/2} = 20(100/i)^{1/2}$$

Degrees Preferential Attachment

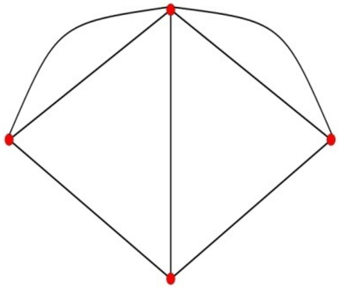


$d_i(t)=m(t/i)^{1/2}$

Nodes with degree < d :

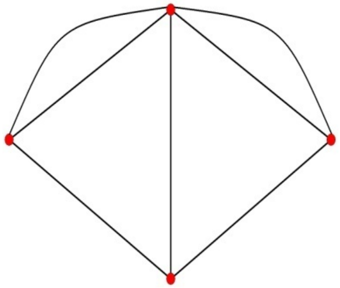
$i > m^2\, t \,/\, d^2$

Distribution of Expected Degrees



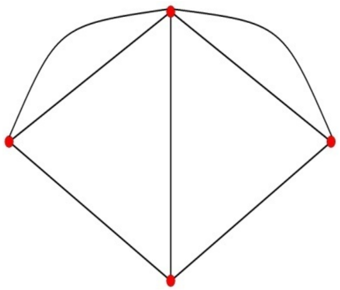
- $dd_i(t)/dt = md_i(t)/2tm = d_i(t)/2t$ and $d_i(i)=m$
- $d_i(t) = m\,(t/i)^{1/2}$
- critical i for some d: $i(d) = m^2\,t\,/\,d^2$

Distribution of Expected Degrees



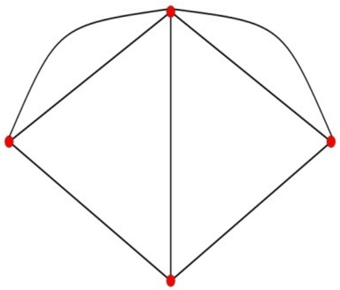
- $dd_i(t)/dt = md_i(t)/2tm = d_i(t)/2t$ and $d_i(i)=m$
- $d_i(t) = m\,(t/i)^{1/2}$
- critical i for some d: $i(d) = m^2 t / d^2$
- $F_t(d) = 1 - i(d)/t = 1 - m^2 / d^2$

Distribution of Expected Degrees



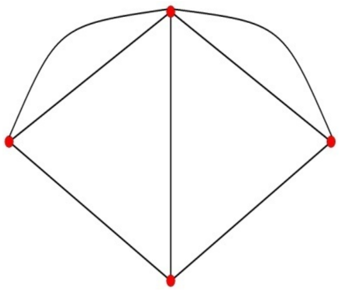
- $dd_i(t)/dt = md_i(t)/2tm = d_i(t)/2t$ and $d_i(i)=m$
- $d_i(t) = m\,(t/i)^{1/2}$
- critical i for some d: $i(d) = m^2 t / d^2$
- $F_t(d) = 1 - i(d)/t = 1 - m^2 / d^2$ so $f_t(d) = 2m^2/d^3$

Power Law



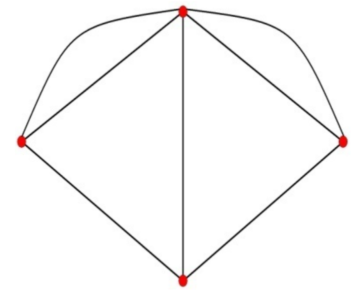
- $f_t(d) = 2m^2/d^3$
- $\log(f(d)) = \log(2m^2) - 3\log(d)$
- Why 3??
- Came from the $dd_i(t)/dt = d_i(t)/2t$

Preferential Attachment?



- Gives power law
- But: why preferential attachment??
- Still not allowing us to do much beyond vary mean degree
- Can we fit degree distributions?

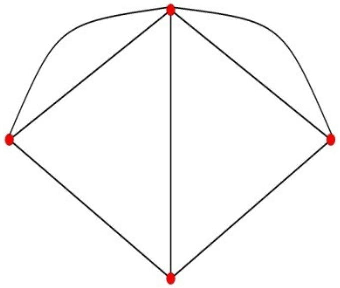
Meeting `Friends of Friends'



JR2007:

- Find new nodes via others: Network-based search
- New node meets *a*m nodes uniformly at random and directs links to them
- Meets (1-*a*)m of their neighbors and attaches to them too

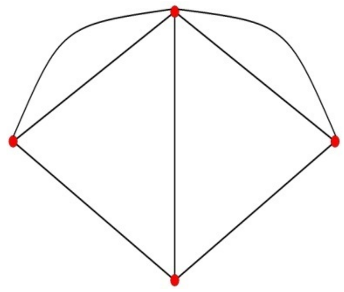
Friends of Friends



- The distribution of neighbors' nodes is not the same as the degree distribution – even with independent link formation

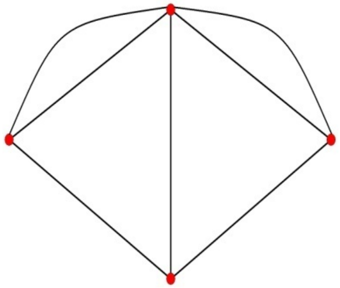
- A neighbor is more likely to be higher degree

Friends of Friends



- Randomly find a node
- Randomly pick one of its neighbors
- Chance of finding some node via this procedure is proportional to its degree: find it if find one of its neighbors....

Simple Hybrid

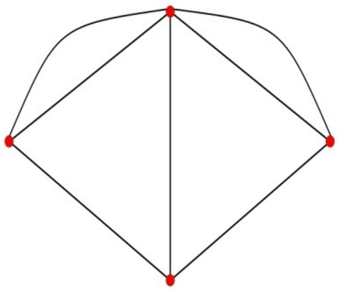


- Fraction *a* uniformly at random,
- Fraction 1-*a* via friends of friends:

$$dd_i(t)/dt = am/t + (1-a)d_i(t)/2t \quad \text{and} \quad d_i(i)=m$$

$$d_i(t) = (m + 2am/(1-a))(t/i)^{(1-a)/2} - 2am/(1-a)$$

Degree distribution



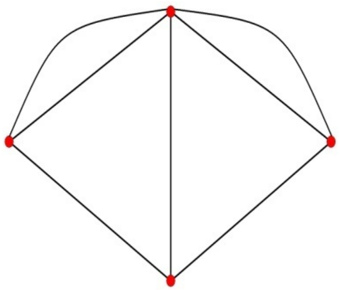
Nodes that have expected degree less than d at some time t are those i such that

$$(m + xam)(t/i)^{1/x} - xam < d \quad \text{where} \quad x = 2/(1-a)$$

critical i is such that

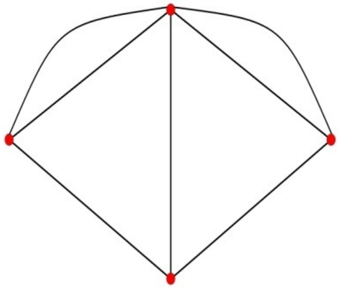
$$i = [(m + xam)/ (d + xam)]^{x}\ t$$

Degree Distribution



- $F(d) = 1 - ((m+amx)/(d+amx))^x \qquad x = 2/(1-a)$
- $a \to 1$ get exponential, $a=0$ get preferential

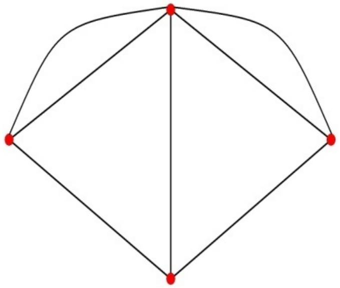
Spans Extremes



- $F(d) = 1 - ((m+amx)/(d+amx))^x$ $\quad x = 2/(1-a)$

- $a \to 1$ get exponential, $a=0$ get preferential

Fitting to data:



- $F(d) = 1 - ((m+amx)/(d+amx))^x \quad x = 2/(1-a)$

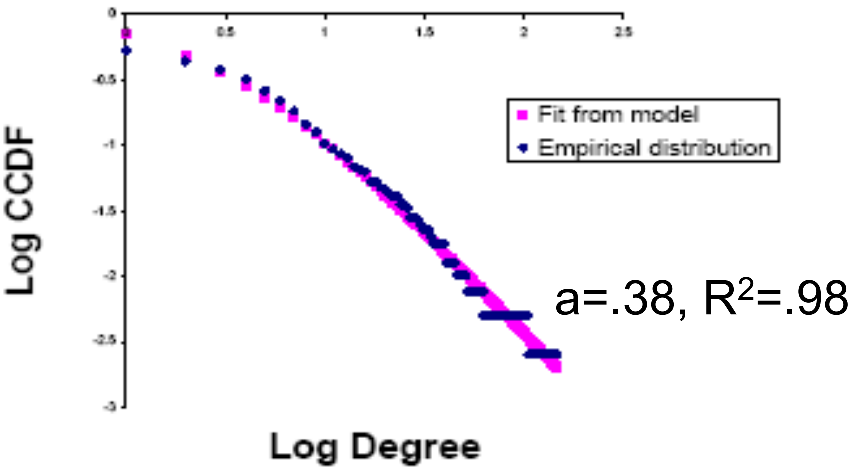
$$\log(1-F(d)) = c - x \log(d+amx)$$

[MLE ?? GMM ??]

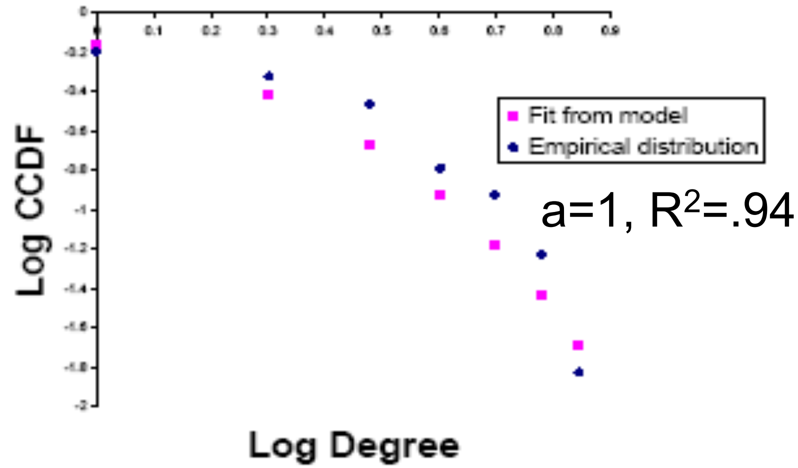
- estimate m directly

- start a_0 gives x_0, then estimate
  $\log(1-F(d)) = c - x_1 \log(d+a_0 m x_0)$ gives a_1 , iterate, look for fixed point

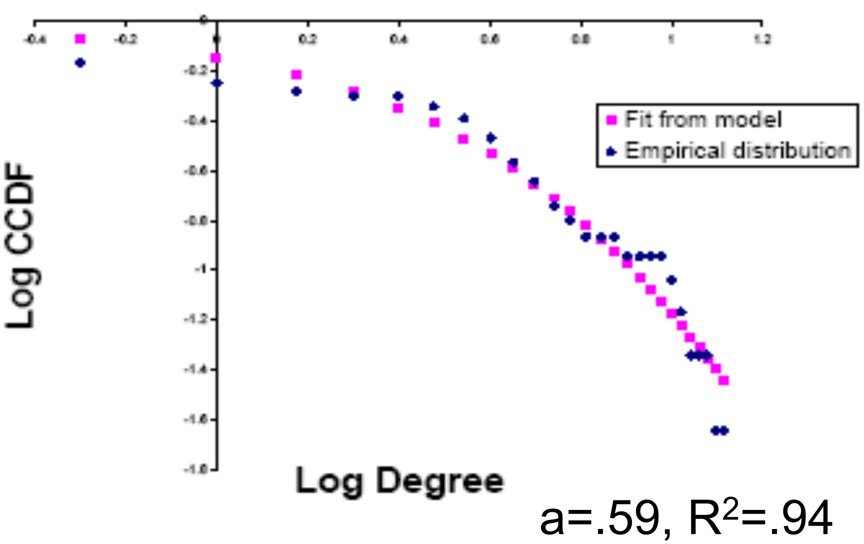
Small World Citations



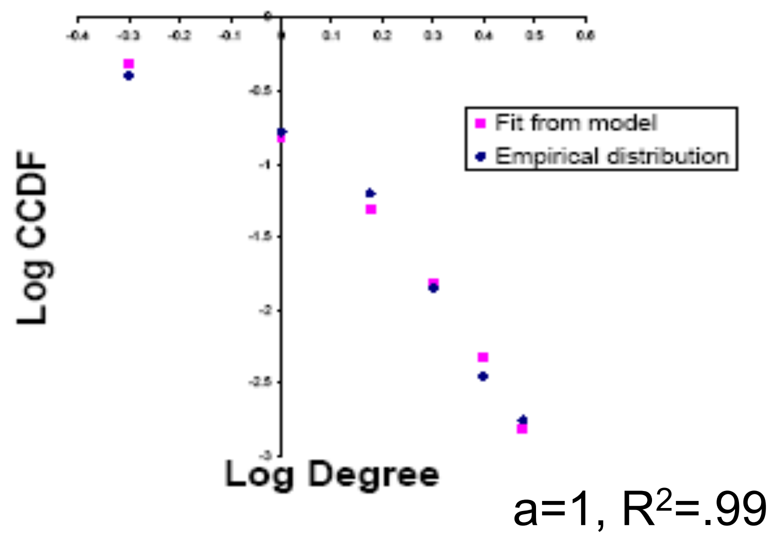
Prison Inmate Friendships



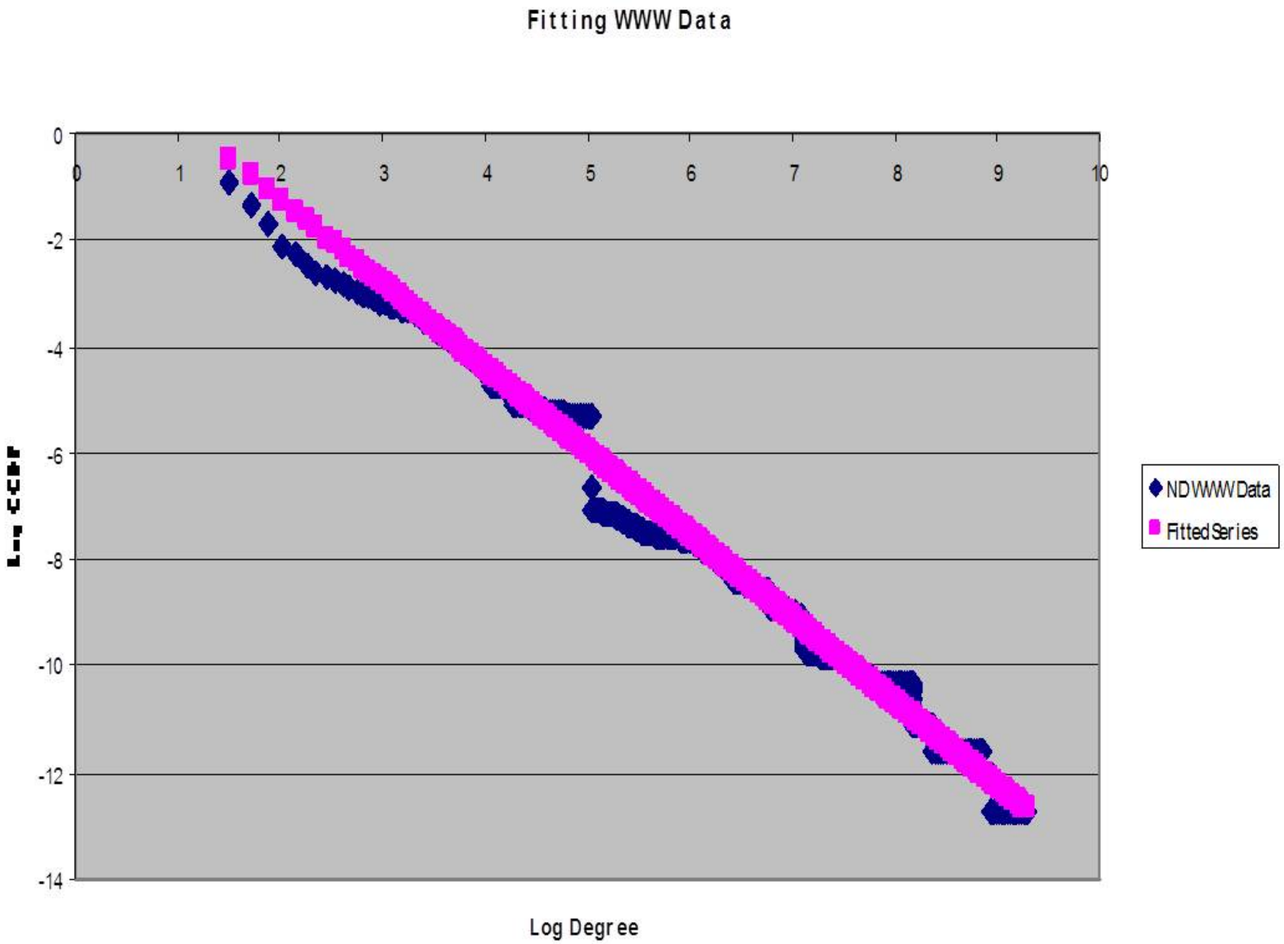
Ham Radio



High School Romance



Degree – ND www Albert, Jeong,



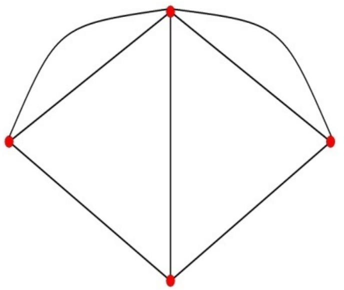
log(freq)

a=.36

R^2=.97

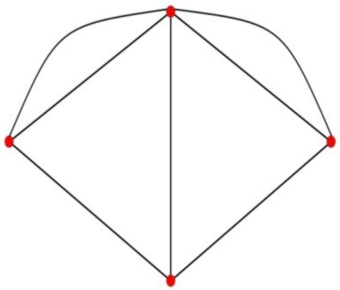
log(degree)

Preferential Attachment?



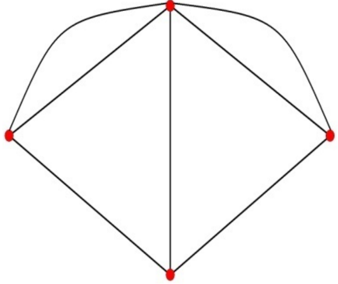
- Fit of Barabasi and Albert has a=.36
- More than 1/3 at random
- Eyeballing Log-Log plots can be misleading!!
- Fat Tails – Yes,   Actual Power law - No

Clustering?



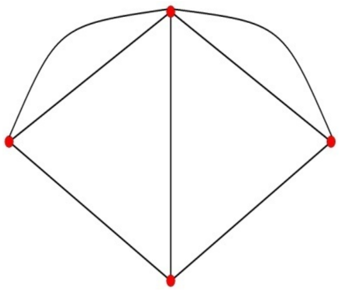
- Yes:
  - connect to friend and friend of friend
  - forms triangles
  - clustering relates to m and a

Approaches



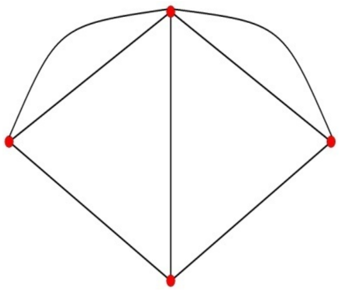
- Classic Random Graphs (how, no why, benchmarks)
- Strategic Formation (why, hard to solve/estimate)
- Growing Random Graphs (some why, but no welfare, limited class, no homophily)
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Approaches



- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

Block model



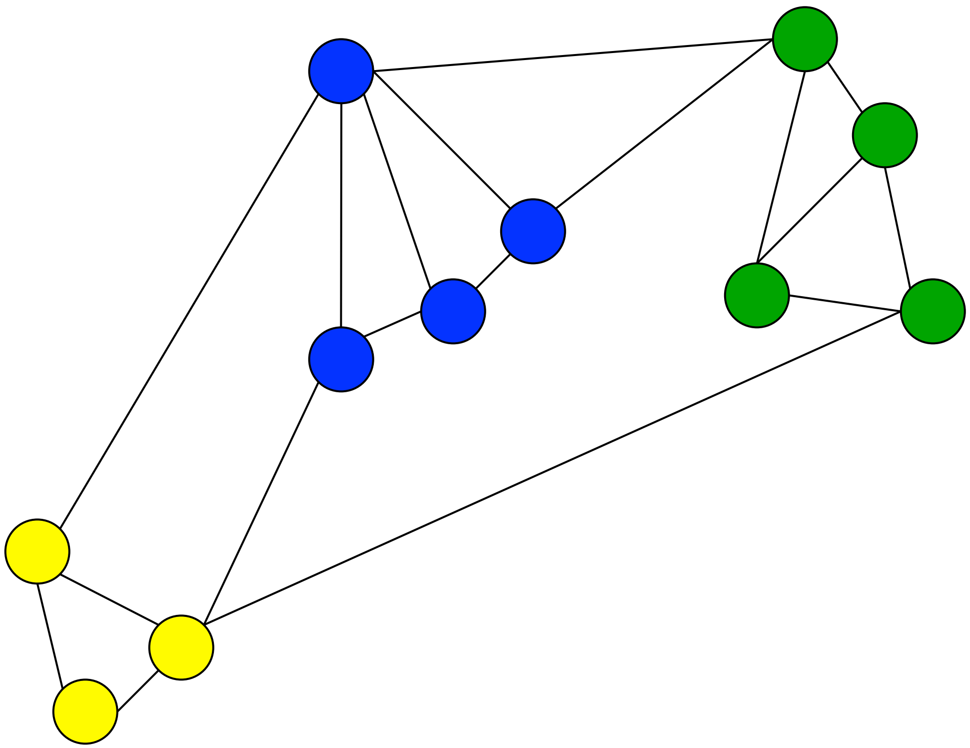
Extend the basic Erdos-Renyi G(n,p) model:

Nodes have characteristics:

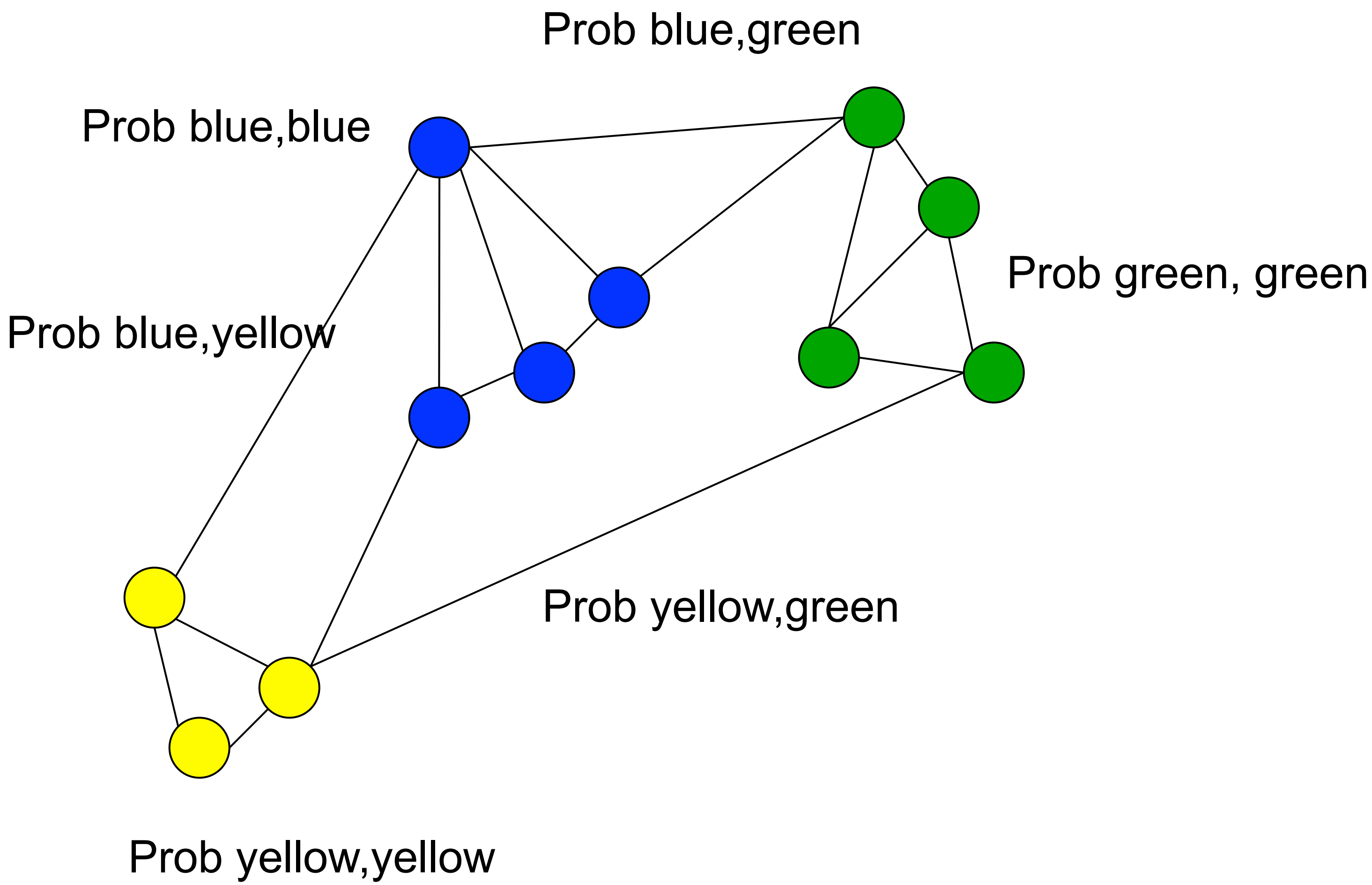
e.g., age, gender, religion, profession, etc.

links between nodes depend on the pairs' characteristics

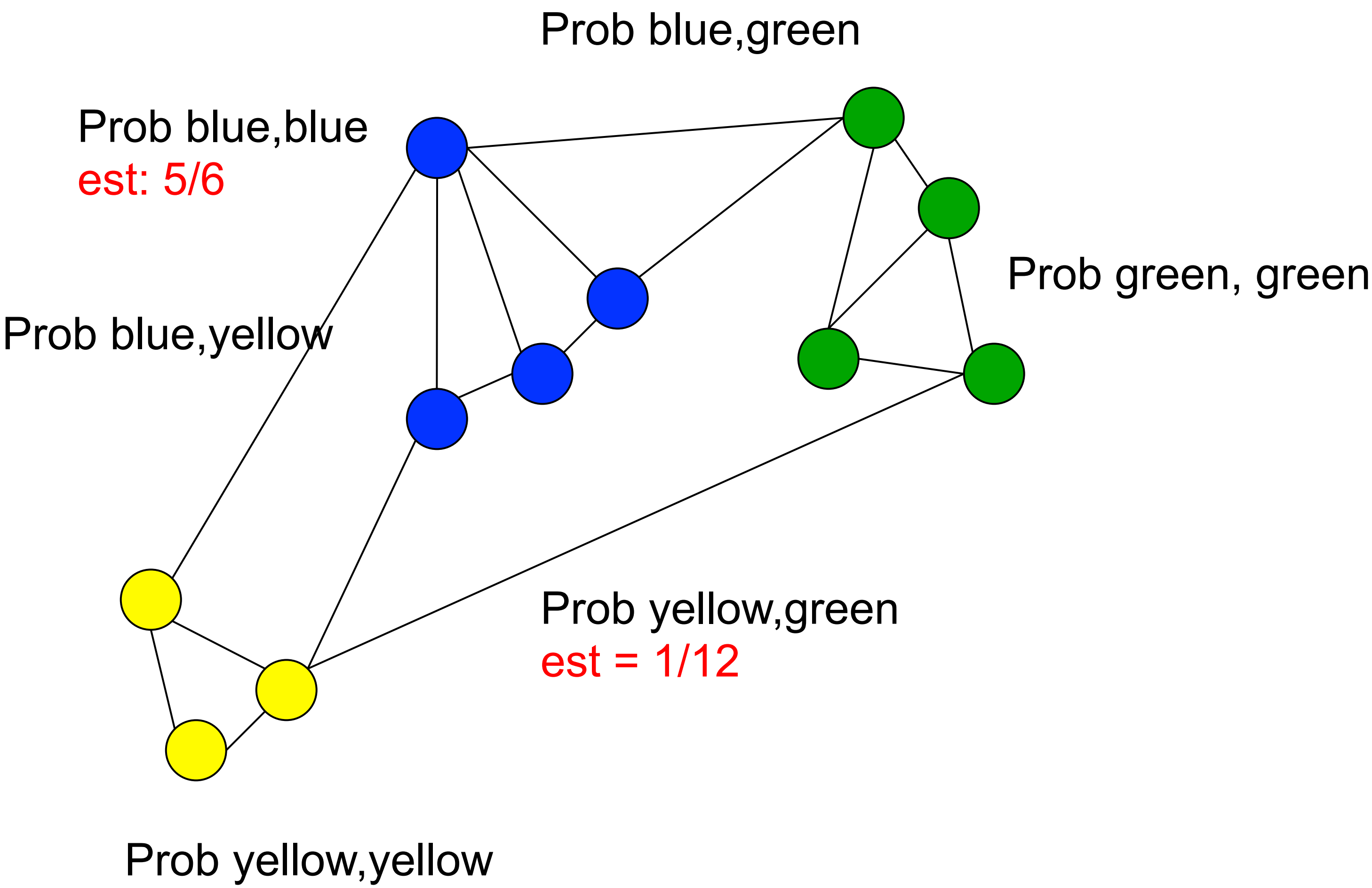
Networks with attributes



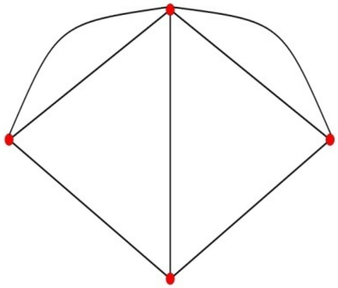
Networks with attributes



Networks with attributes



Example:



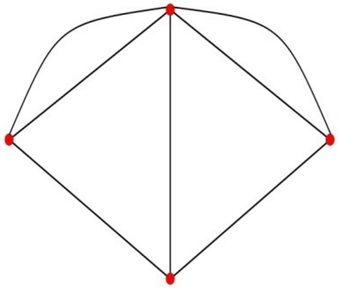
- $\{1, \ldots, n\}$ agents/nodes

- Partitioned into groups $N_1, \ldots, N_K$

- Node i in group k is linked to a node j in group k' with probability $P_{kk'}$ (undirected)

- Homophily: $P_{kk} > P_{kk'}$ for $k' \neq k$

Block models



Continuous covariates:

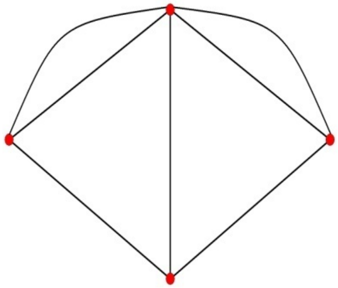
Example: link between i and j depends on their characteristics:

$$\beta_i X_i + \beta_j X_j + \beta_{ij} |X_i - X_j|$$

E.g.,

$$\mathrm{Log}(p_{ij} / (1-p_{ij})) = \beta_i X_i + \beta_j X_j + \beta_{ij} |X_i - X_j|$$

Block models

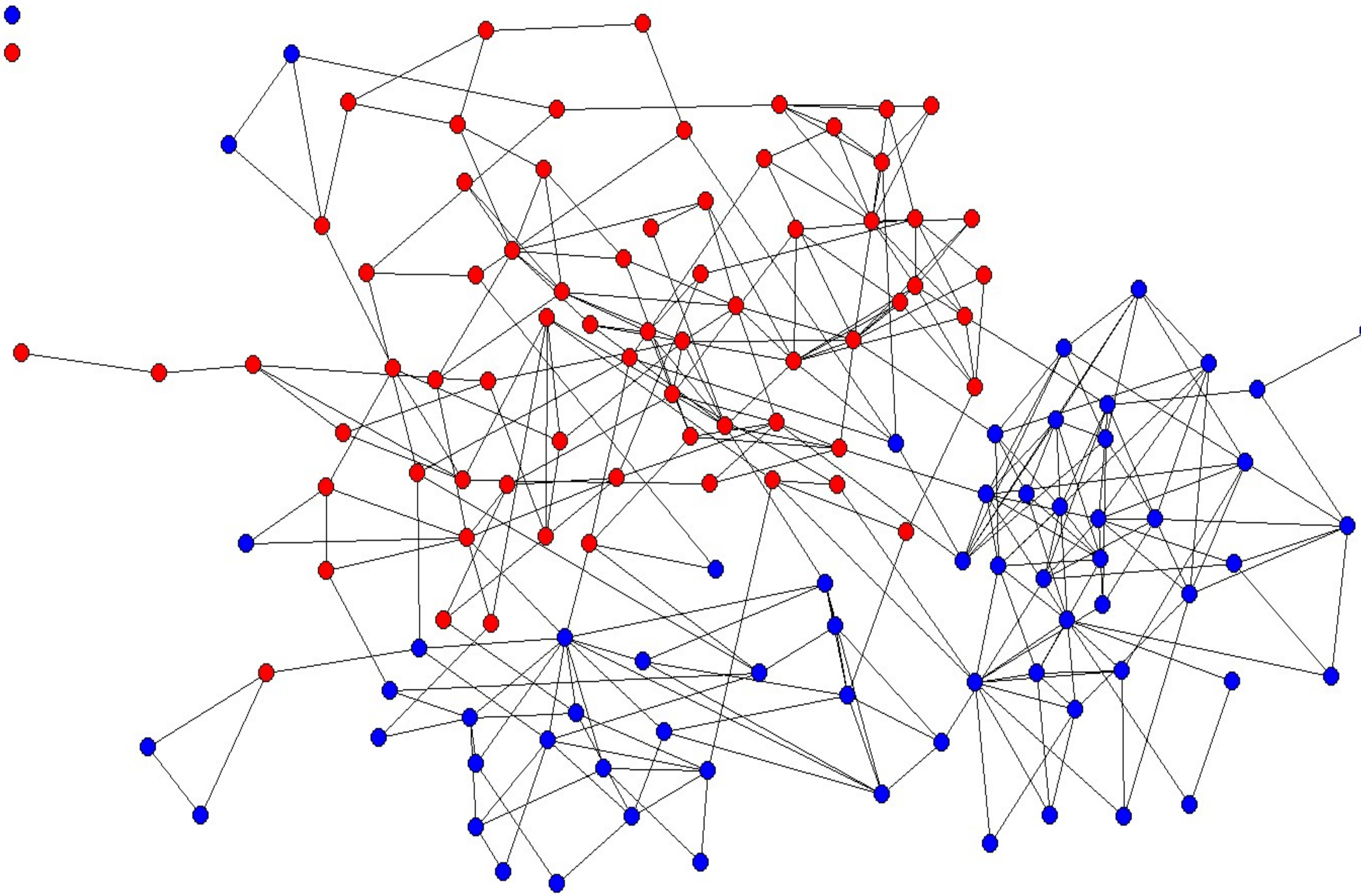


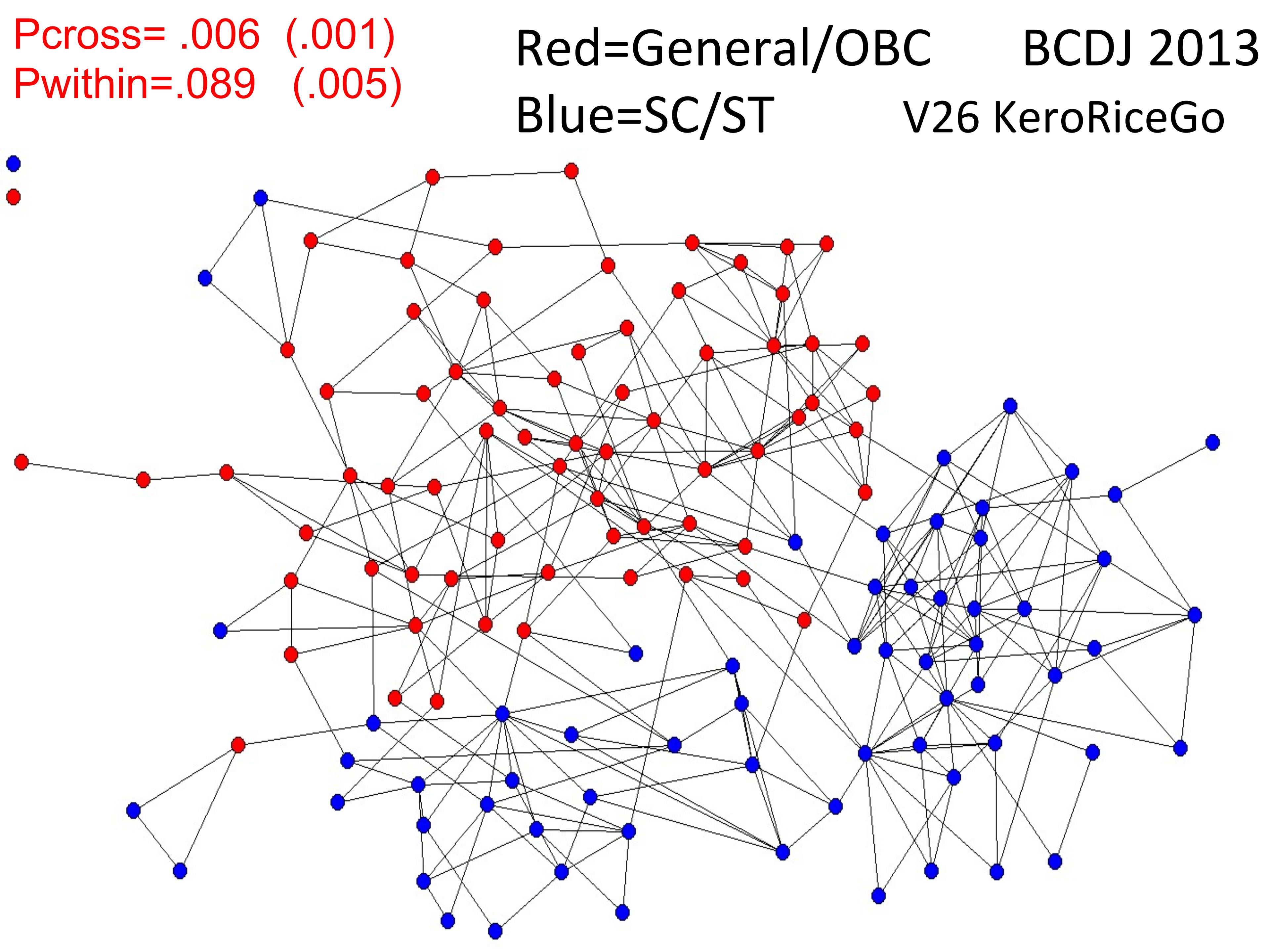
Could use this sort of model

to test for homophily...

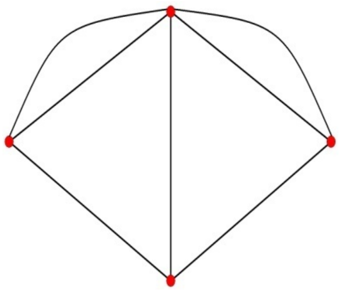
Red=General/OBC BCDJ 2013

Blue=SC/ST V26 KeroRiceGo



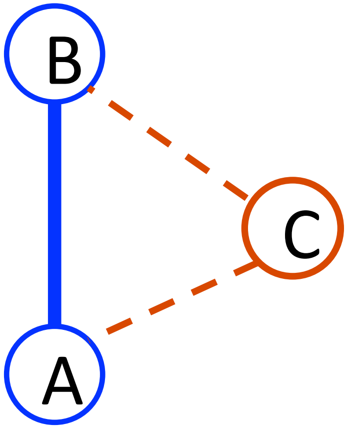

Pcross= .006 (.001)
Pwithin=.089 (.005)
Red=General/OBC
Blue=SC/ST
BCDJ 2013
V26 KeroRiceGo

What is missed?



- Likelihood of link depends on node attributes (observed or latent)

- *also depends on whether nodes have friends in common*

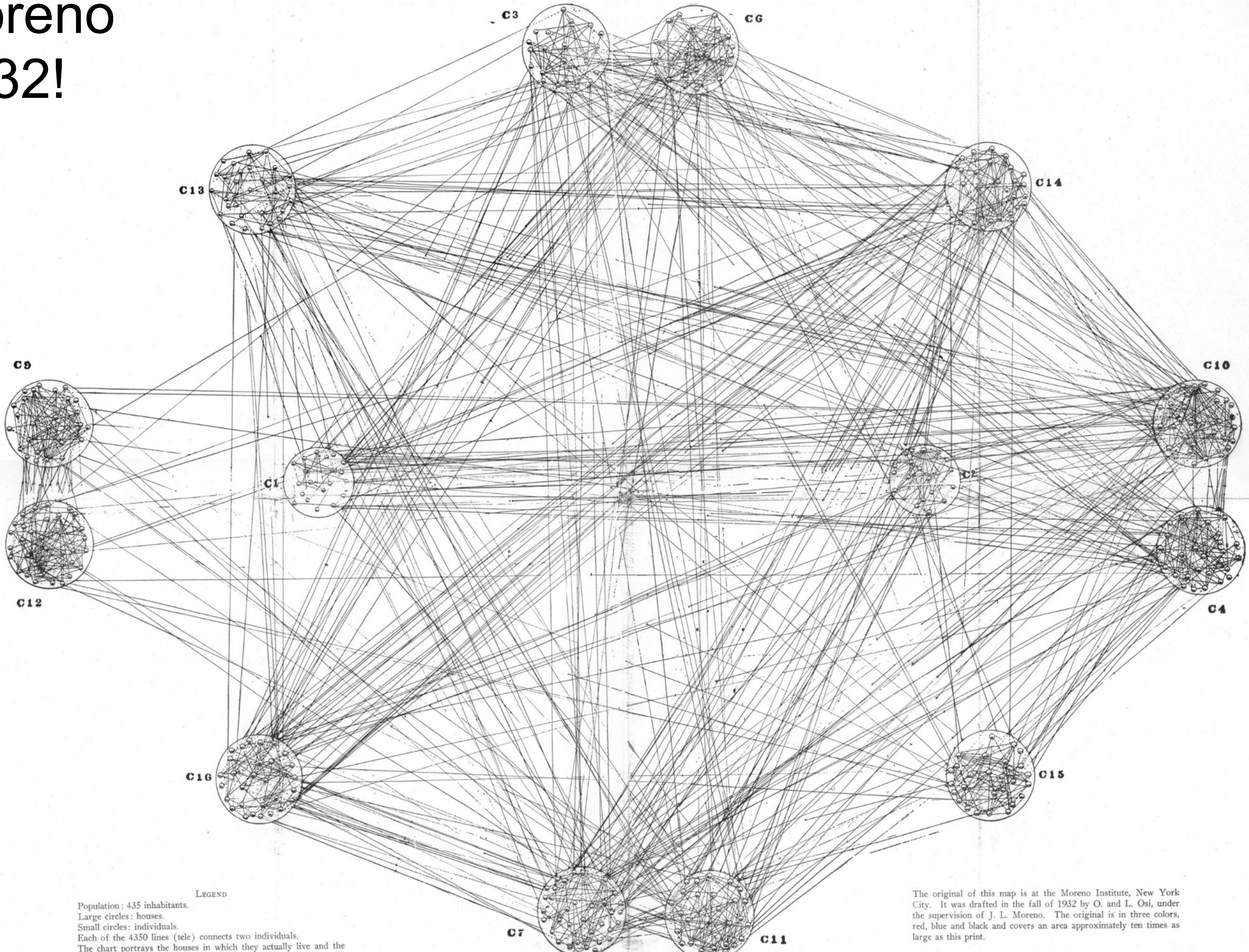


“A pertinent form of statistical treatment would be one which deals with social configurations as wholes, and not with single series of facts, more or less artificially separated from the total picture.”

Jacob Levy Moreno and Helen Hall Jennings, 1938.

Moreno
1932!

SOCIOMETRIC GEOGRAPHY OF A COMMUNITY – MAP III

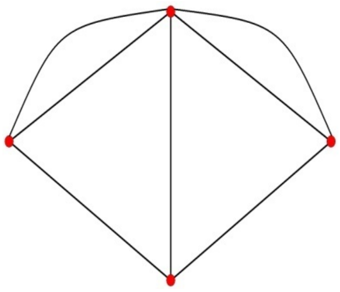


LEGEND

Population: 435 inhabitants.
Large circles: houses.
Small circles: individuals.
Each of the 4350 lines (tele) connects two individuals.
The chart portrays the houses in which they actually live and the houses in which they would like to live.

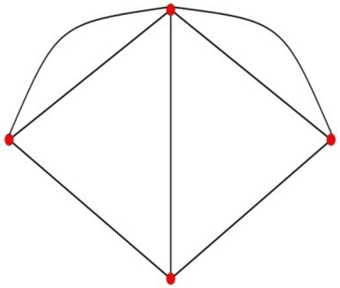
The original of this map is at the Moreno Institute, New York City. It was drafted in the fall of 1932 by O. and L. Osi, under the supervision of J. L. Moreno. The original is in three colors, red, blue and black and covers an area approximately ten times as large as this print.

Approaches



- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models
  - ERGMS
  - SUGMS
  - Implications of Pairwise stability
  - Approximate independence with distance

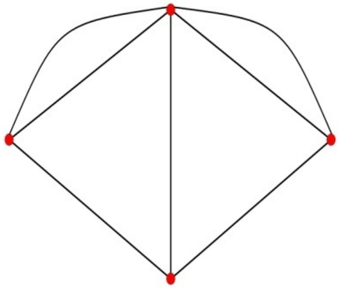
ERGMs



Example: probability depends on

$$\beta_L \text{ \#links}(g) + \beta_T \text{ \#triangles}(g)$$

ERGMs



Want probability *of network* to depend on

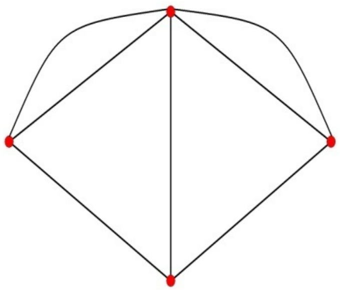
$$\beta_L\, L(g) + \beta_T\, T(g)$$

Set

$$\color{red}{Pr(g)} \sim \exp[\, \beta_L\, L(g) + \beta_T\, T(g)\,]$$

(now positive)

ERGMs



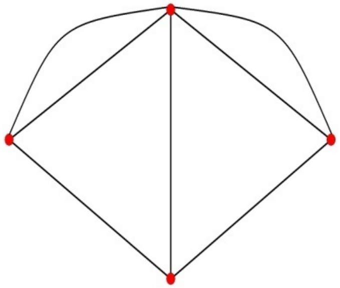
Want probability to depend on

$$\beta_L \, L(g) + \beta_T \, T(g)$$

Set $\Pr(g) \sim \exp[\, \beta_L \, L(g) + \beta_T \, T(g) \,]$

Theorem by Hammersly and Clifford (71): *any* network model can be expressed in the exponential family with counts of graph statistics

Example: Erdos-Renyi G(n,p)



- p – probability of a link, L(g) - number of links in g

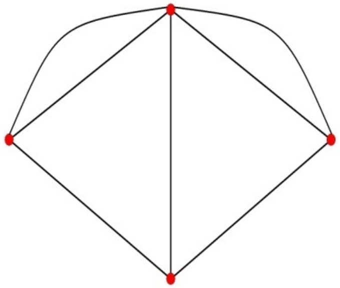
$$\Pr[(g)] = p^{L(g)}(1-p)^{n(n-1)/2-L(g)}$$

$$= [p/(1-p)]^{L(g)}(1-p)^{n(n-1)/2}$$

$$= \exp[\log(p/(1-p))\, L(g) - \log(1/(1-p))n(n-1)/2\,]$$

$$= \exp[\quad \beta_1 \quad s_1(g) \quad - \quad c \quad]$$

ERGMs

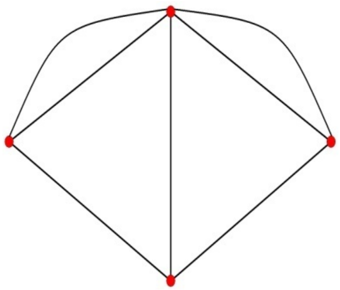


To be probability:

$$\Pr(g) = \frac{\exp[\ \beta_L\ L(g) + \beta_T\ T(g)\ ]}{\sum_{g'} \exp[\ \beta_L\ L(g') + \beta_T\ T(g')]}$$

$$\Pr(g) = \exp[\ \beta_L\ L(g) + \beta_T\ T(g) - c]$$

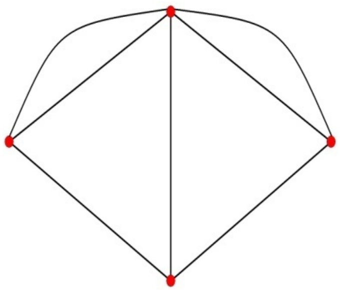
ERGMs



$$\Pr(g) = \frac{\exp[\ \beta_1 s_1(g) + \ldots + \beta_k s_k(g)\ ]}{\sum_{g'} \exp[\ \beta_1 s_1(g') + \ldots + \beta_k s_k(g')\ ]}$$

- MCMC techniques for estimation (Snijders 02, Handcock 03,...) have led to these becoming the standard

Issues:



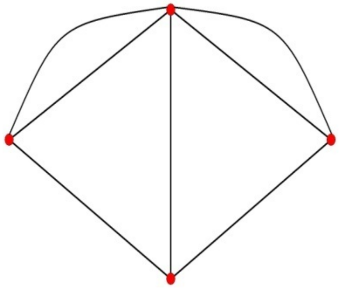
- $\Pr(g) = \dfrac{\exp[\,\beta_1 s_1(g) + \ldots + \beta_k s_k(g)\,]}{\sum_{g'} \exp[\,\beta_1 s_1(g') + \ldots + \beta_k s_k(g')\,]}$

- Recall: n=30 nodes, there are 2^{435} g's

(less than 2^{275} atoms in the universe...)

- ***Sampling g's will not lead to accurate estimates (not just MCMC limitation)***

ERGMs



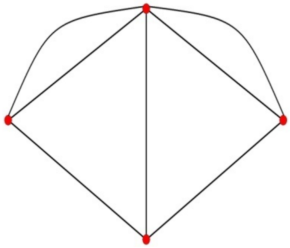
Bhamidi, Bresler and Sly (2008)
(see also Chatterjee and Diaconis (2011)):

For dense enough ERGMs, MCMC (Glauber dynamics - Gibbs sampling) estimates mix less than exponentially ***only if*** networks have approximately independent links

So, ERGMs that are interesting, cannot be estimated via techniques being used!

Simulations: also problems on sparse ones...

Example:

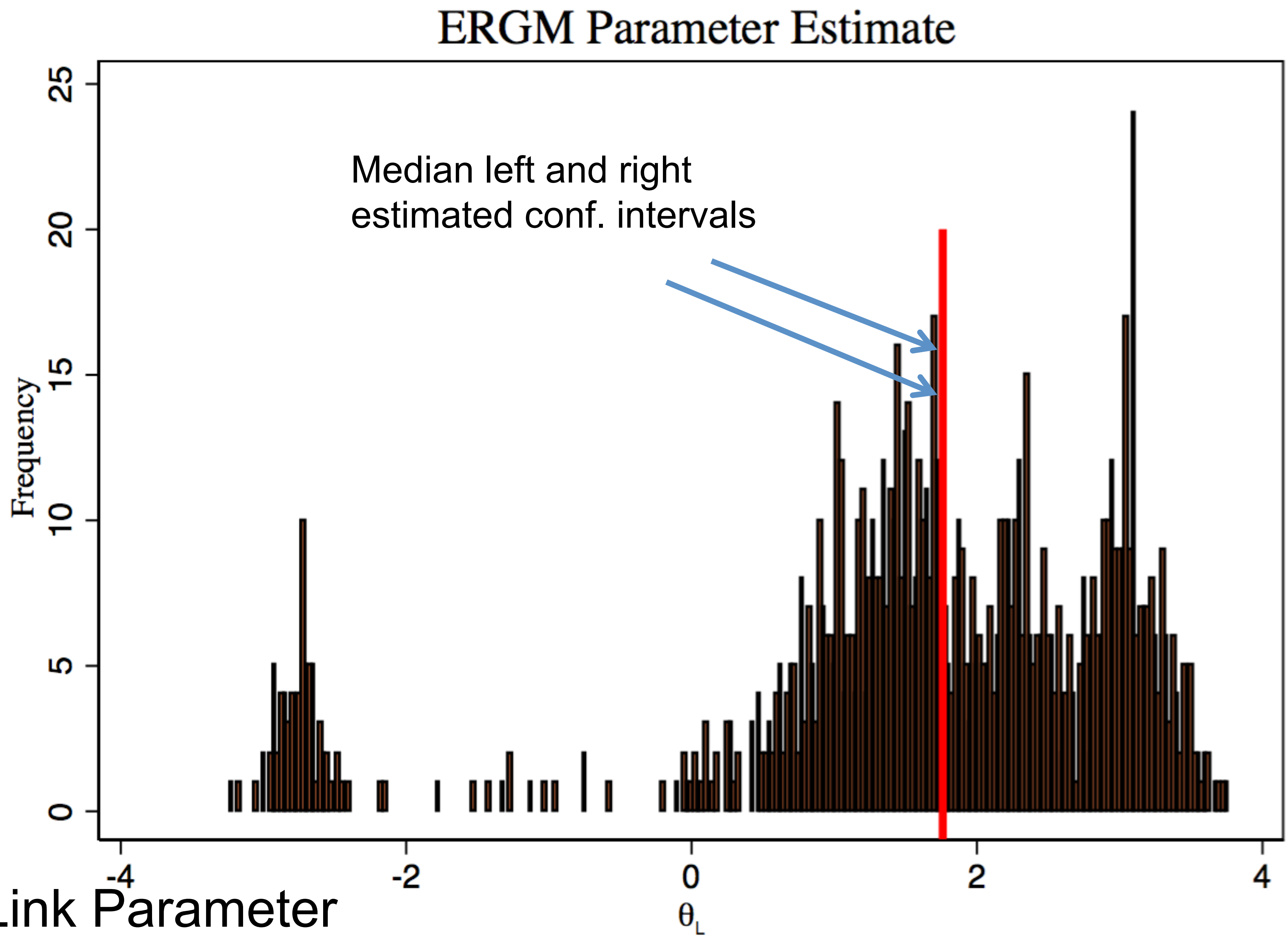


- $\Pr(g) = \dfrac{\exp[\beta_I I(g) + \beta_L L(g) + \beta_T T(g)]}{\sum_{g'} \exp[\beta_I I(g) + \beta_L L(g') + \beta_T T(g')]}$

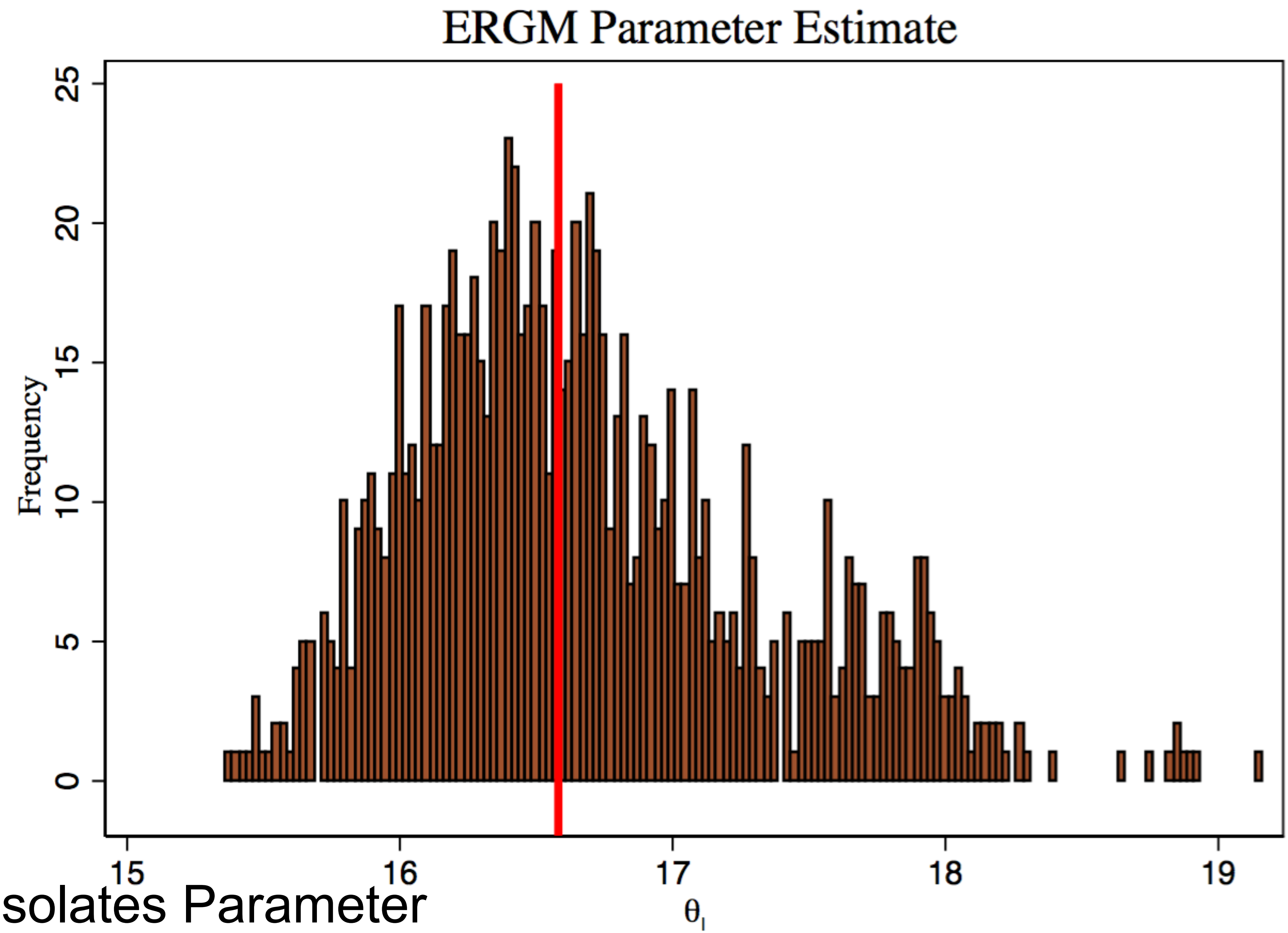
- I(g) = #isolates(g)
- L(g) = #links(g)
- T(g) = #triangles(g)

- n=50 nodes, 1000 estimations of networks based on same statistics:

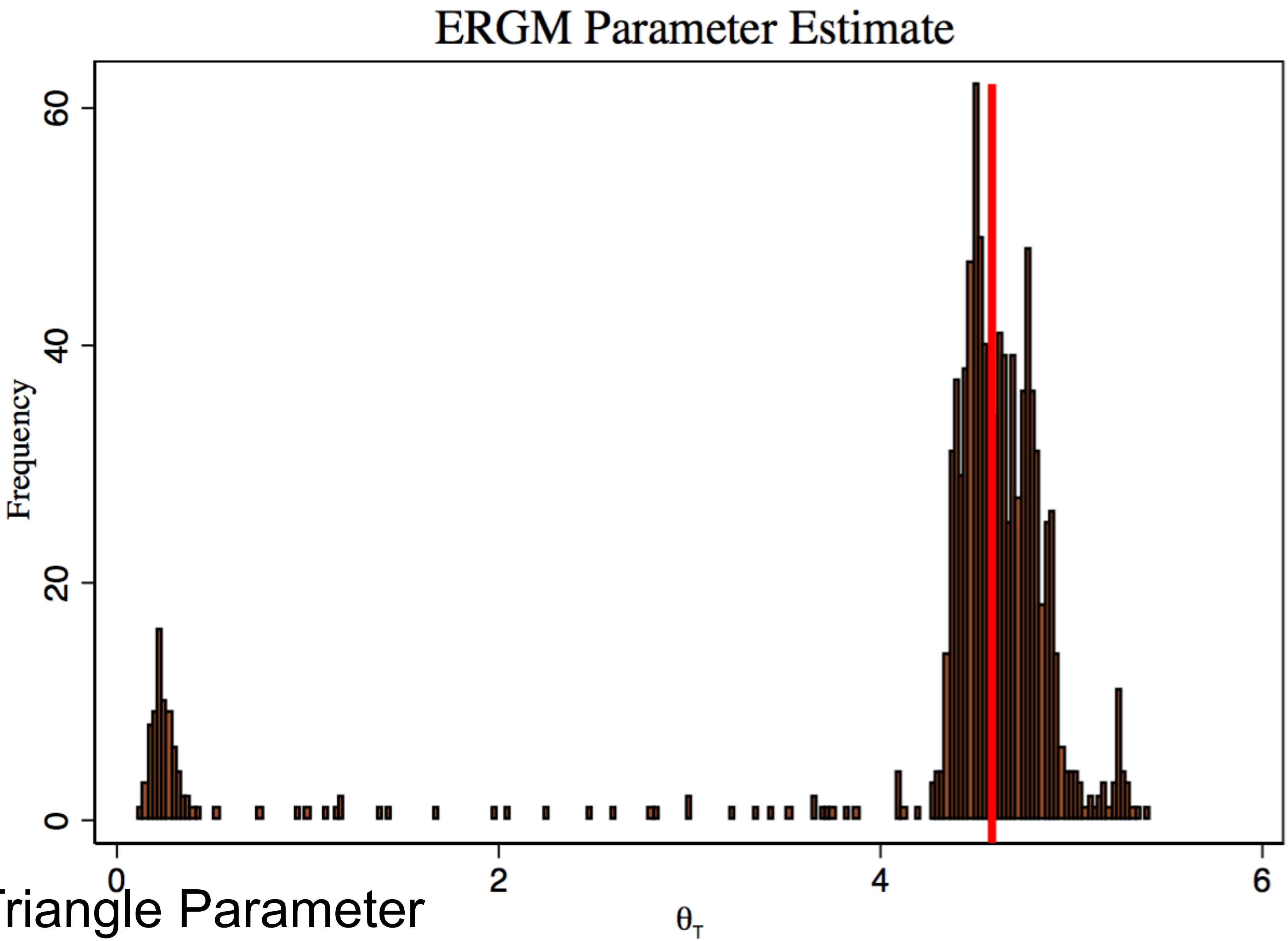
  20 isol, 10 triangles, 45 links



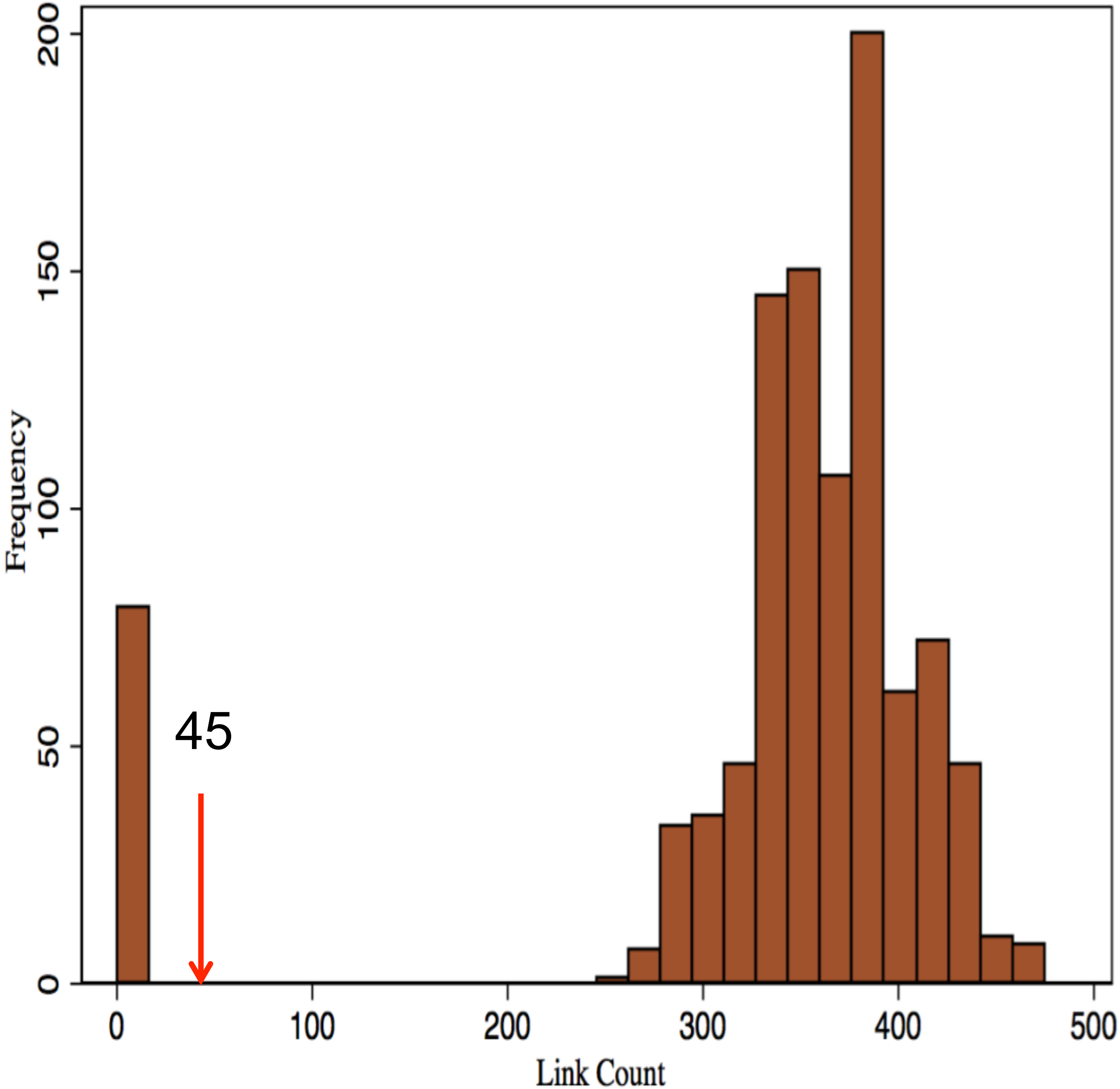
Link Parameter



Isolates Parameter

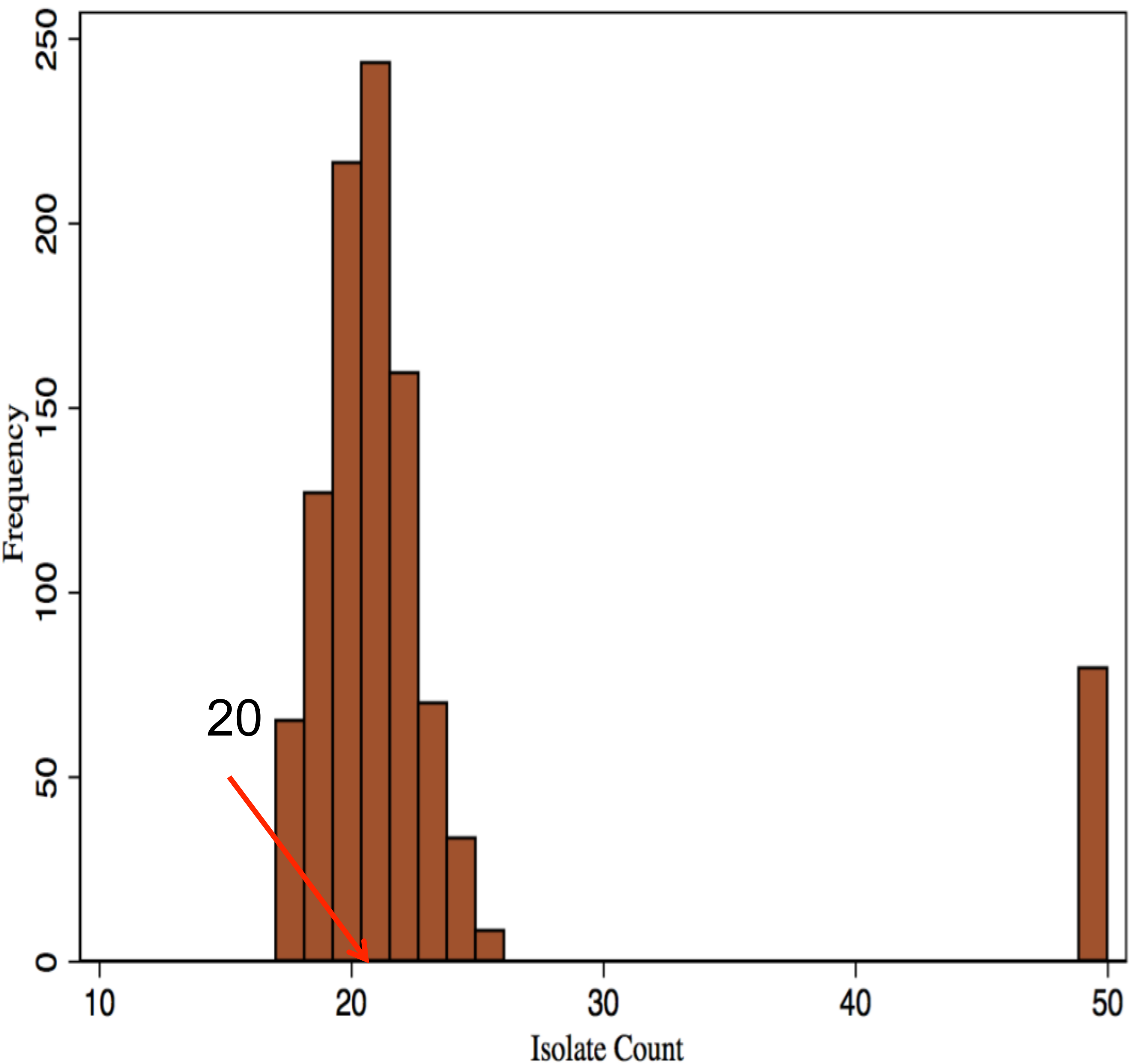


Triangle Parameter

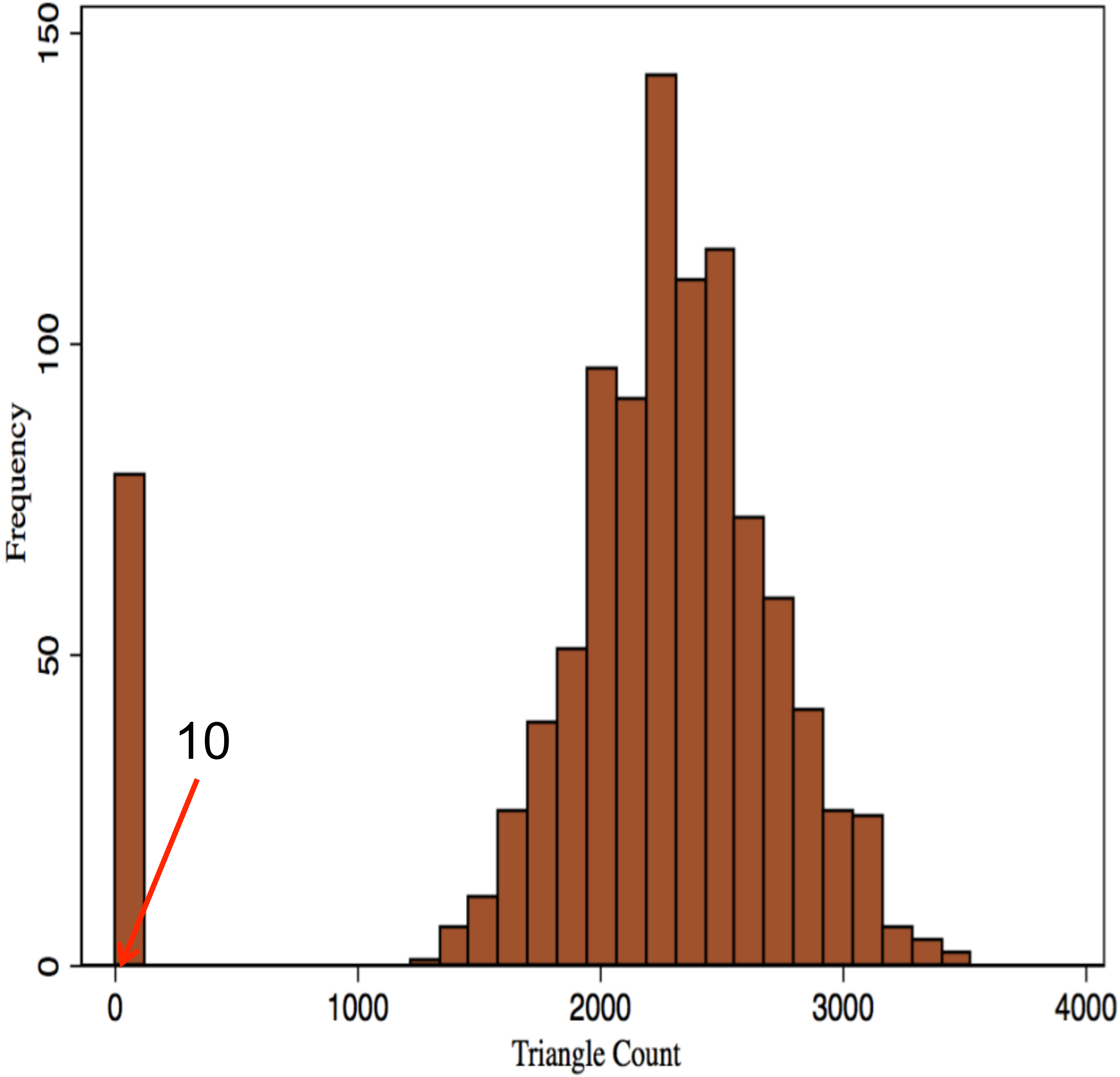


Recreate Links

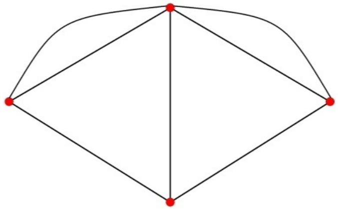
Recreate Isolates



Recreate Triangles

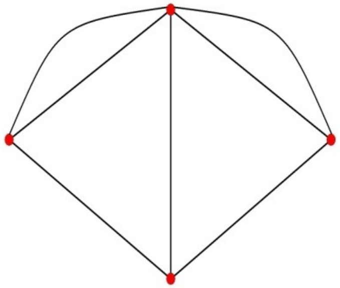


Issues:



- MCMC estimation techniques are inaccurate:
  - Can one compute parameters?
- Consistency of estimators of ERGMs:
  - When are parameters accurate and how many nodes are needed?
- How to generate networks randomly?
  - Counterfactuals, validation…

Approaches



- Classic Random Graphs
- Strategic Formation
- Growing Random Graphs
- Econometric models
  - Block models (only links, no externalities...)
  - ERGMS (not estimable, which formulation?...)
  - SUGMS (estimable, but only with subgraphs)
  - Implications of Pairwise stability (multiple equilibria...)
  - Approximate independence with distance (not valid in many settings...)