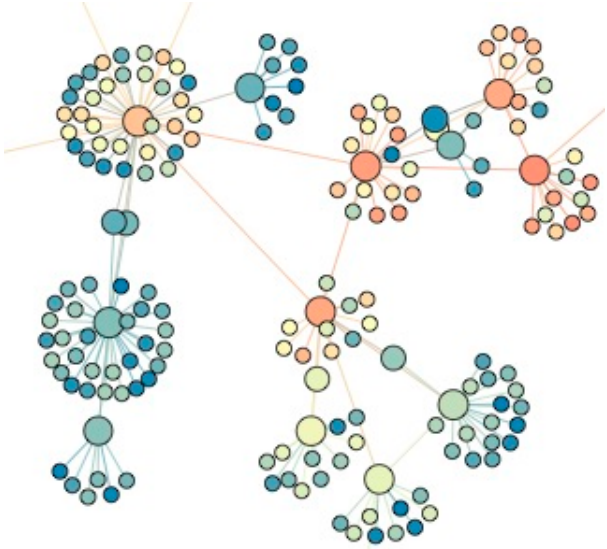


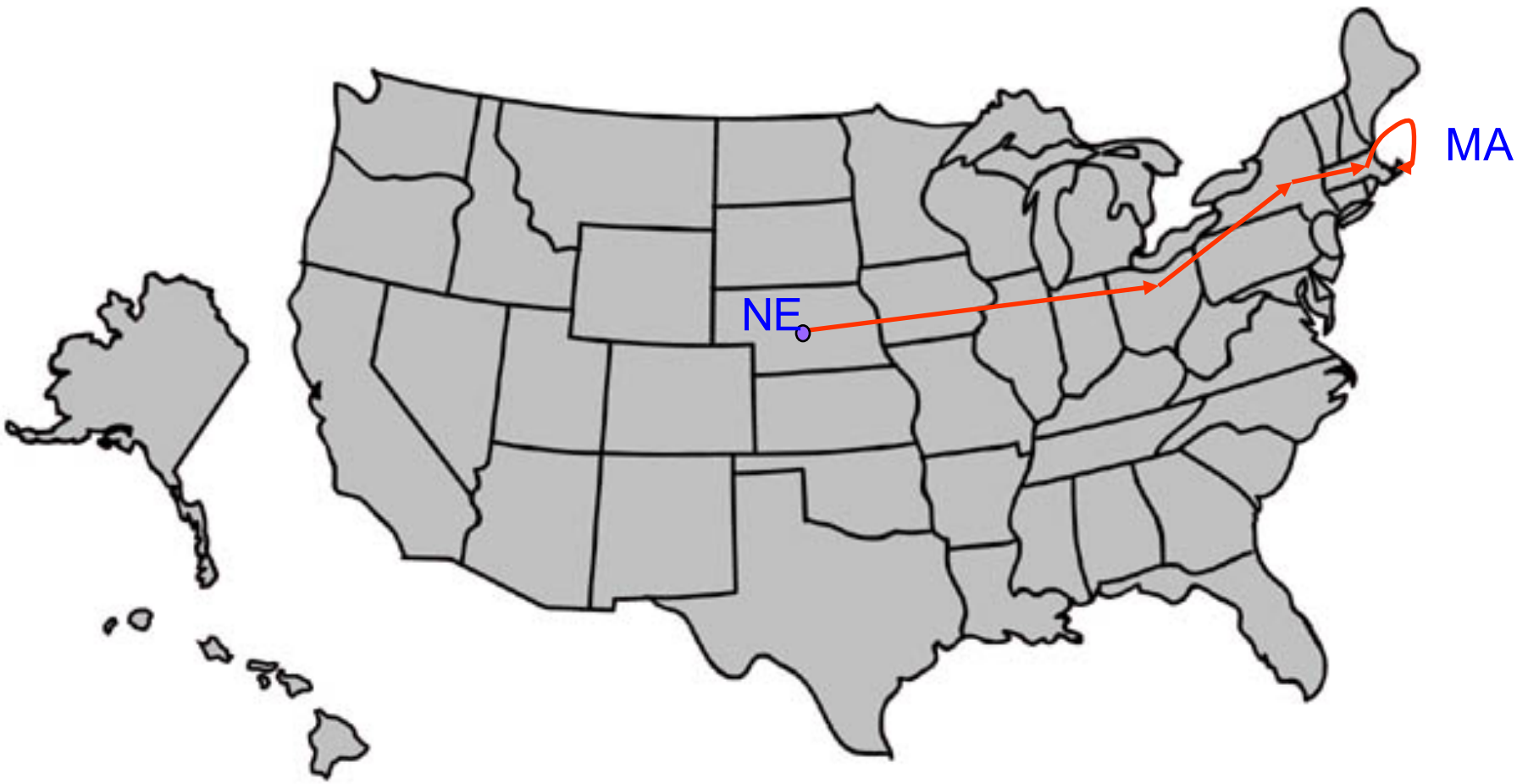
Small world networks

CS 224W

Outline

- Small world phenomenon
  - Milgram's small world experiment
- Local structure
  - clustering coefficient
  - motifs
- Small world network models:
  - Watts & Strogatz (clustering & short paths)
  - Kleinberg (geographical)
  - Kleinberg, Watts/Dodds/Newman (hierarchical)
- Small world networks: why do they arise?

Small world phenomenon:
Milgram's experiment



Milgram's experiment

Instructions:
Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target.

Outcome:

20% of initiated chains reached target
average chain length = 6.5

- "Six degrees of separation"

Milgram's experiment repeated

email experiment
Dodds, Muhamad, Watts,
Science 301, (2003)
(optional reading)

- 18 targets
- 13 different countries

- 60,000+ participants
- 24,163 message chains
- 384 reached their targets
- average path length 4.0

Source: NASA, U.S. Government; http://visibleearth.nasa.gov/view_rec.php?id=2429

Interpreting Milgram's experiment

- Is 6 is a ***surprising*** number?
  - In the 1960s? Today? Why?

- Pool and Kochen in (1978 established that the average person has between 500 and 1500 acquaintances)

Quiz Q:

- Ignore for the time being the fact that many of your friends' friends are your friends as well. If everyone has 500 friends, the average person would have how many friends of friends?
  - 500
  - 1,000
  - 5,000
  - 250,000

Quiz Q:

- With an average degree of 500, a node in a random network would have this many friends-of-friends-of-friends (3rd degree neighbors):
  - 5,000
  - 500,000
  - 1,000,000
  - 125,000,000

Interpreting Milgram's experiment

- Is 6 is a ***surprising*** number?
  - In the 1960s? Today? Why?
- If social networks were random… ?
  - Pool and Kochen (1978) - ~500-1500 acquaintances/person
  - ~ 500 choices 1st link
  - ~ 500^2 = 250,000 potential 2nd degree neighbors
  - ~ 500^3 = 125,000,000 potential 3rd degree neighbors
- If networks are completely cliquish?
  - all my friends' friends are my friends
  - what would happen?

Quiz Q:

- If the network were completely cliquish, that is all of your friends of friends were also directly your friends, what would be true:
  - (a) None of your friendship edges would be part of a triangle (closed triad)
  - (b) It would be impossible to reach any node outside the clique by following directed edges
  - (c) Your shortest path to your friends' friends would be 2

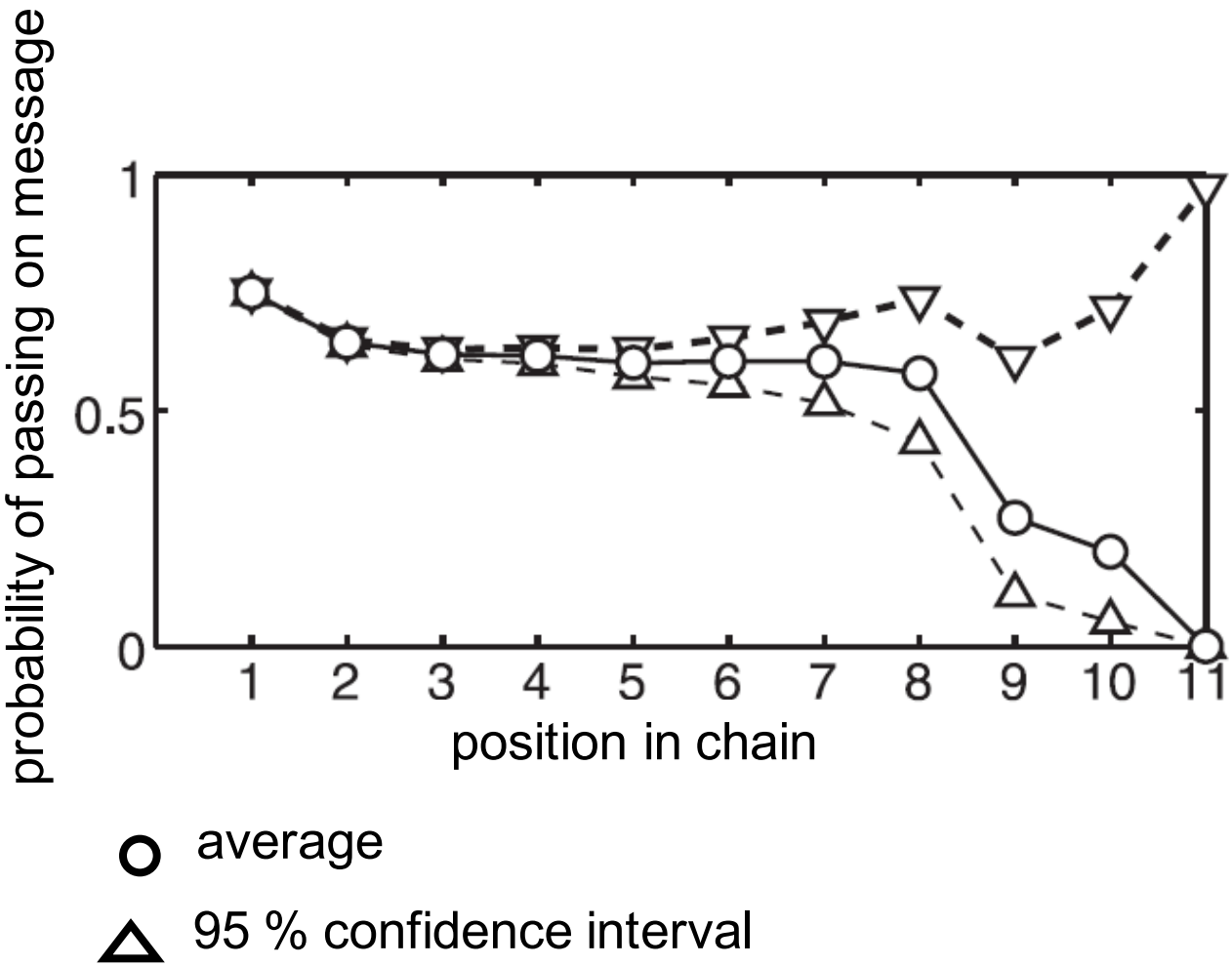
complete cliquishness

- If all your friends of friends were also your friends, you would be part of an isolated clique.

Uncompleted chains and distance

- Is 6 an ***accurate*** number?
- What bias is introduced by uncompleted chains?
  - are longer or shorter chains more likely to be completed?

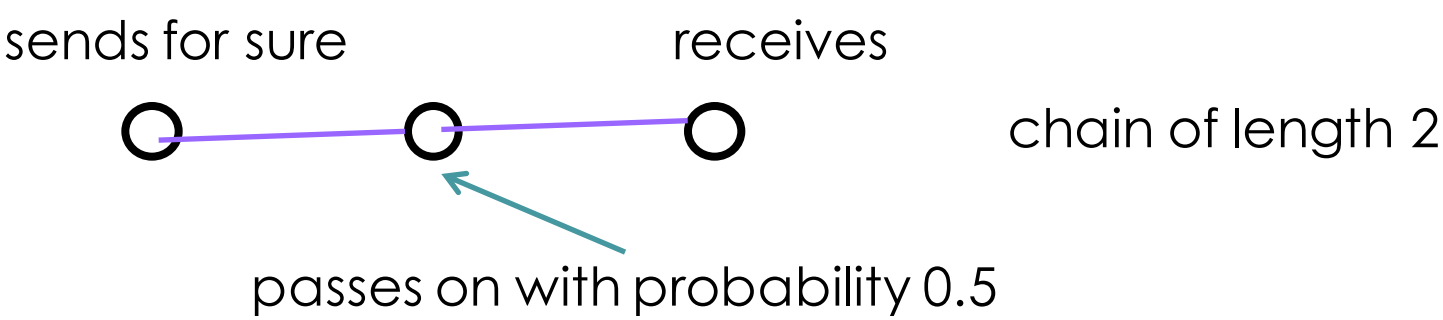
Attrition



Source: An Experimental Study of Search in Global Social Networks: Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts (8 August 2003); Science 301 (5634), 827.

Quiz Q:

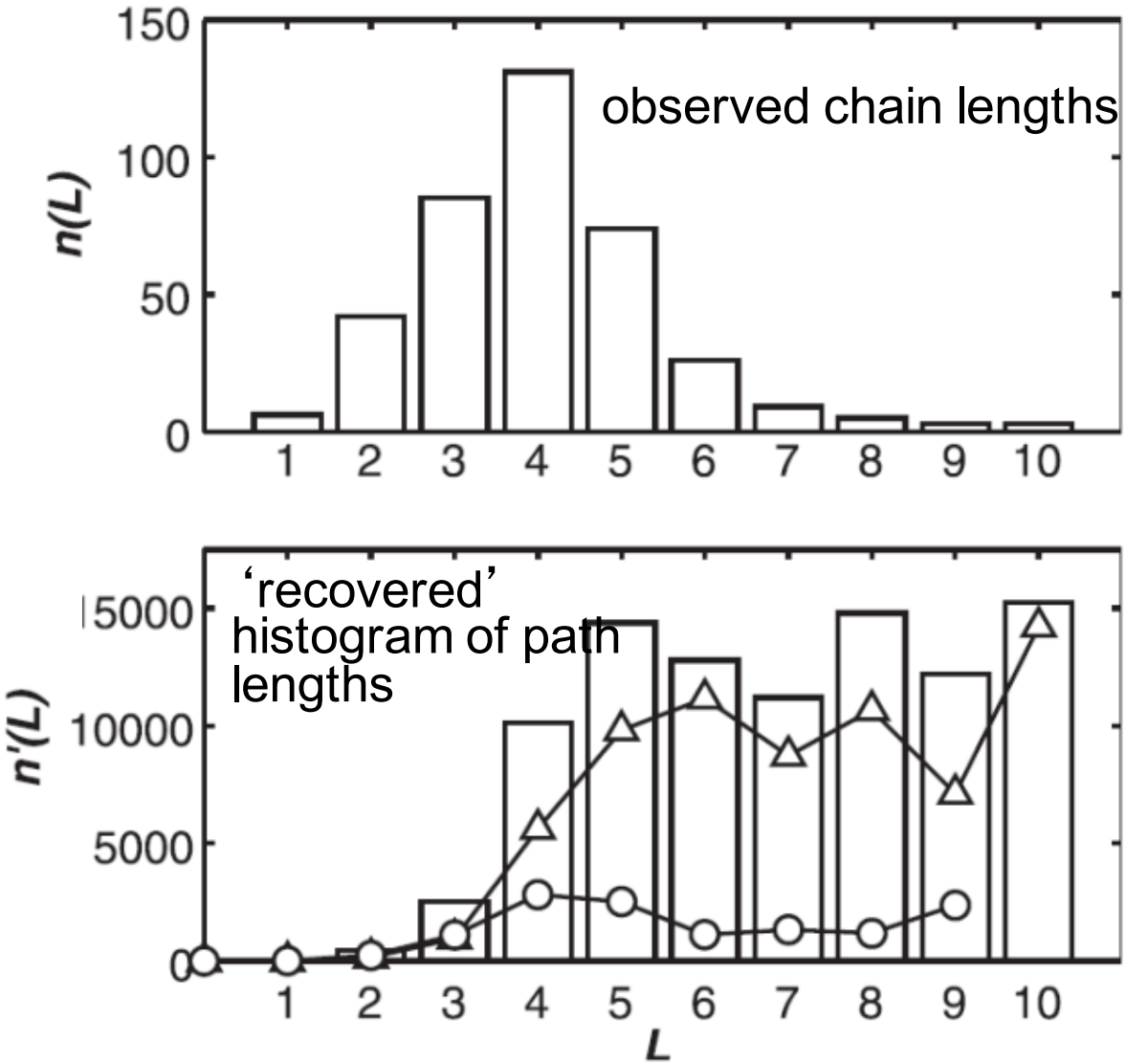
- if each intermediate person in the chain has 0.5 probability of passing the letter on, what is the likelihood of a chain being completed
  - of length 2?
  - of length 5?



Quiz Q:

- if each intermediate person in the chain has 0.5 probability of passing the letter on, what is the likelihood of a chain of length 5 being completed
  - (a) ½
  - (b) ¼
  - (c) 1/8
  - (d) 1/16

Estimating the true distance



inter-country

intra-country

Source: An Experimental Study of Search in Global Social Networks: Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts (8 August 2003); Science 301 (5634), 827.

Navigation and accuracy

- Is 6 an ***accurate*** number?

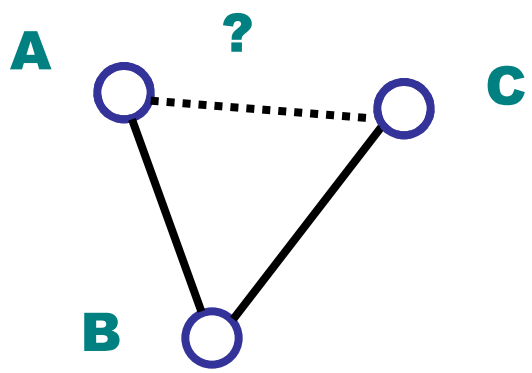
- Do people find the ***shortest*** paths?
  - Killworth, McCarty ,Bernard, & House (2005):
  - less than optimal choice for next link in chain is made ½ of the time

Small worlds & networking

What does it mean to be 1, 2, 3 hops apart on Facebook, Twitter, LinkedIn, Google Plus?

Transitivity, triadic closure, clustering

- Transitivity:
  - if A is connected to B and B is connected to C what is the probability that A is connected to C?
  - my friends' friends are likely to be my friends



Clustering

- Global clustering coefficient

  3 x number of triangles in the graph

  number of connected triples of vertices

$$C = \frac{\text{3 x number of triangles in the graph}}{\text{number of connected triples}}$$

Local clustering coefficient (Watts&Strogatz 1998)

- For a vertex i
  - The fraction pairs of neighbors of the node that are themselves connected
  - Let n_i be the number of neighbors of vertex *i*

$$C_i = \frac{\text{\# of connections between i's neighbors}}{\text{max \# of possible connections between i's neighbors}}$$

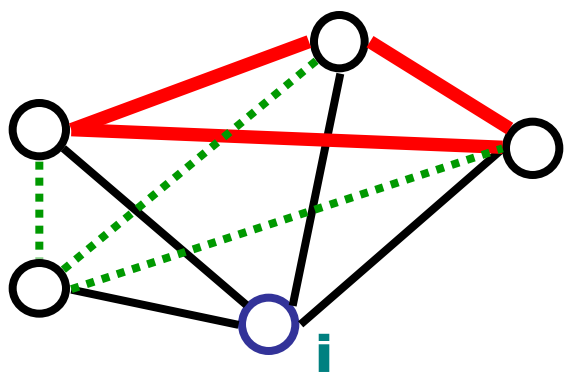
$$C_{i\ \text{directed}} = \frac{\text{\# directed connections between i's neighbors}}{n_i * (n_i - 1)}$$

$$C_{i\ \text{undirected}} = \frac{\text{\# undirected connections between i's neighbors}}{n_i * (n_i - 1)/2}$$

Local clustering coefficient (Watts&Strogatz 1998)

- Average over all n vertices

$$C = \frac{1}{n}\sum_i C_i$$



link present

link absent

$n_i = 4$

max number of connections:

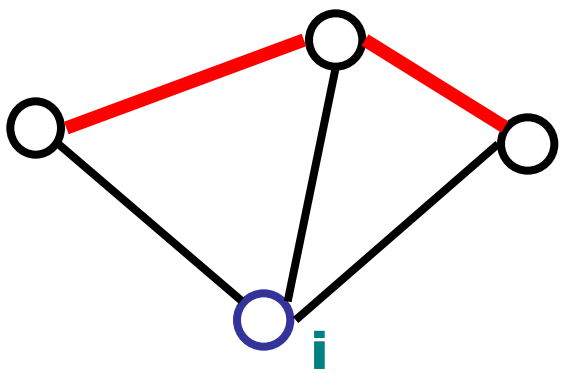
4*3/2 = 6

3 connections present

$C_i = 3/6 = 0.5$

Quiz Q:

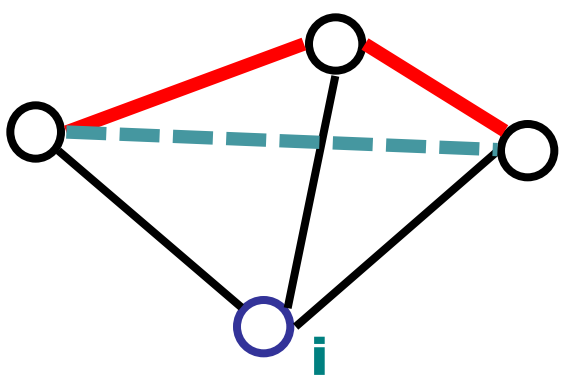
- The clustering coefficient for vertex i is:



(a) 0
(b) 1/3
(c) 1/2
(d) 2/3

Explanation

- $n_i = 3$
- there are 2 connections present out of max of 3 possible
- $C_i = 2/3$



beyond social networks

Small world phenomenon:

high clustering $C_{\text{network}} >> C_{\text{random graph}}$

low average shortest path $l_{\text{network}} \approx \ln(N)$

- what other networks can you think of with these characteristics?

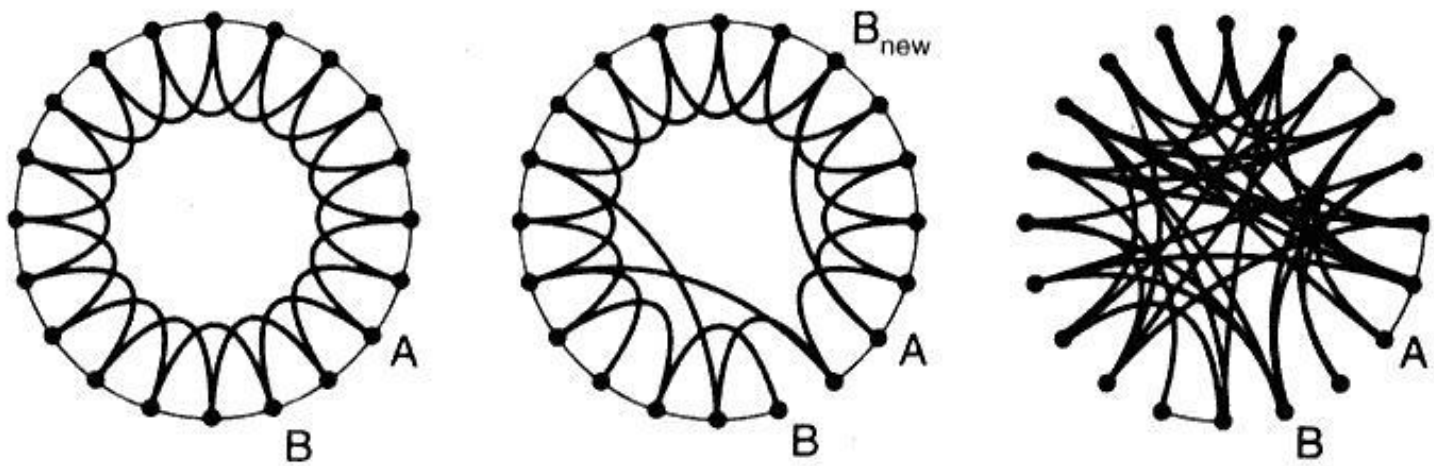
Comparison with "random graph" used to determine whether real-world network is "small world"

Network	size	av. shortest path	Shortest path in fitted random graph	Clustering (averaged over vertices)	Clustering in random graph
Film actors	225,226	3.65	2.99	0.79	0.00027
MEDLINE co-authorship	1,520,251	4.6	4.91	0.56	1.8×10^{-4}
E.Coli substrate graph	282	2.9	3.04	0.32	0.026
C.Elegans	282	2.65	2.25	0.28	0.05

Small world phenomenon:
Watts/Strogatz model

Reconciling two observations:

- **High clustering:** my friends' friends tend to be my friends
- **Short average paths**



Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Watts-Strogatz model:
Generating small world graphs

Select a fraction p of edges
Reposition on of their endpoints

Add a fraction p of additional edges leaving underlying lattice intact

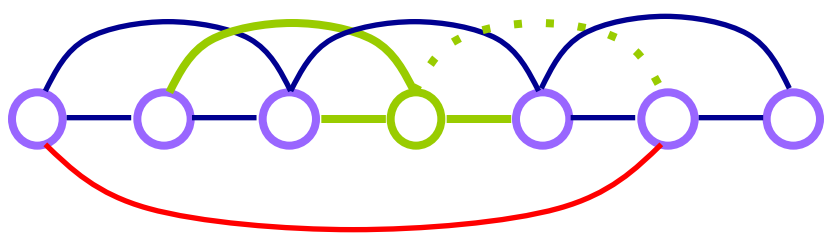
- As in many network generating algorithms
  - Disallow self-edges
  - Disallow multiple edges

Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Watts-Strogatz model:

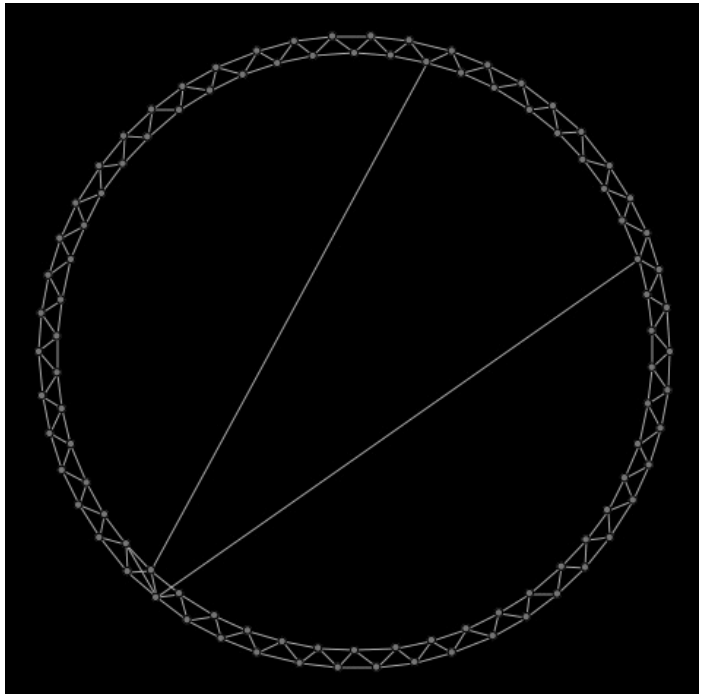
Generating small world graphs

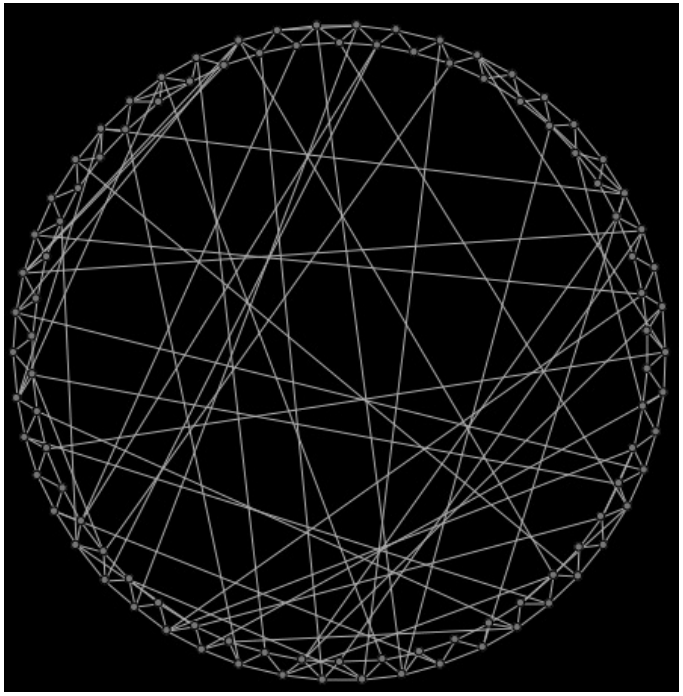
- Each node has K>=4 nearest neighbors (local)
- tunable: vary the probability *p* of rewiring any given edge
- small *p*: regular lattice
- large *p*: classical random graph



Quiz question:

- Which of the following is a result of a higher rewiring probability?



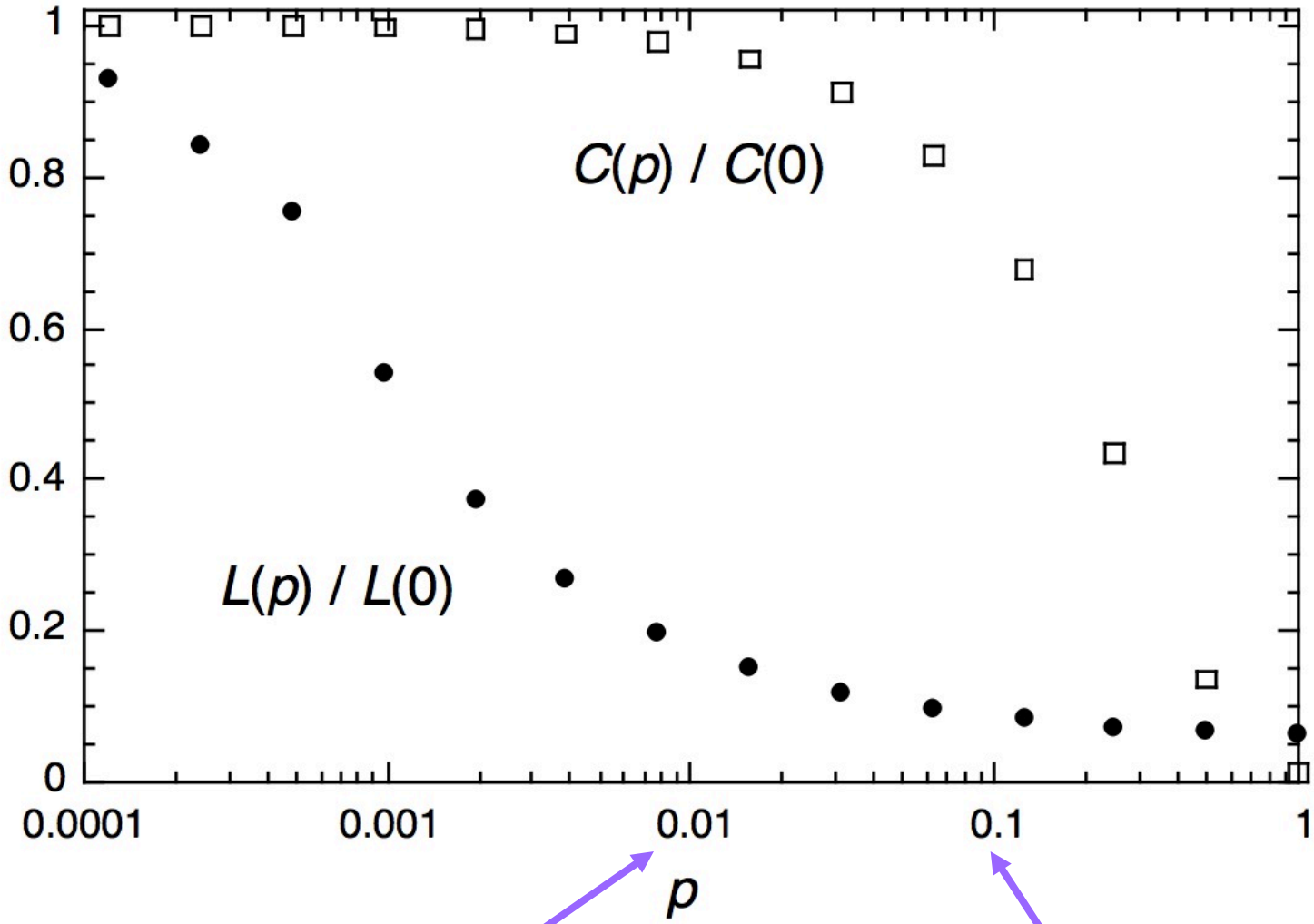


(a) Left (b) Right (c) insufficient information

What happens in between?

- Small shortest path means low clustering?
- Large shortest path means high clustering?
- Through numerical simulation
  - As we increase p from 0 to 1
    - Fast decrease of mean distance
    - Slow decrease in clustering

Clust coeff. and ASP as rewiring increases



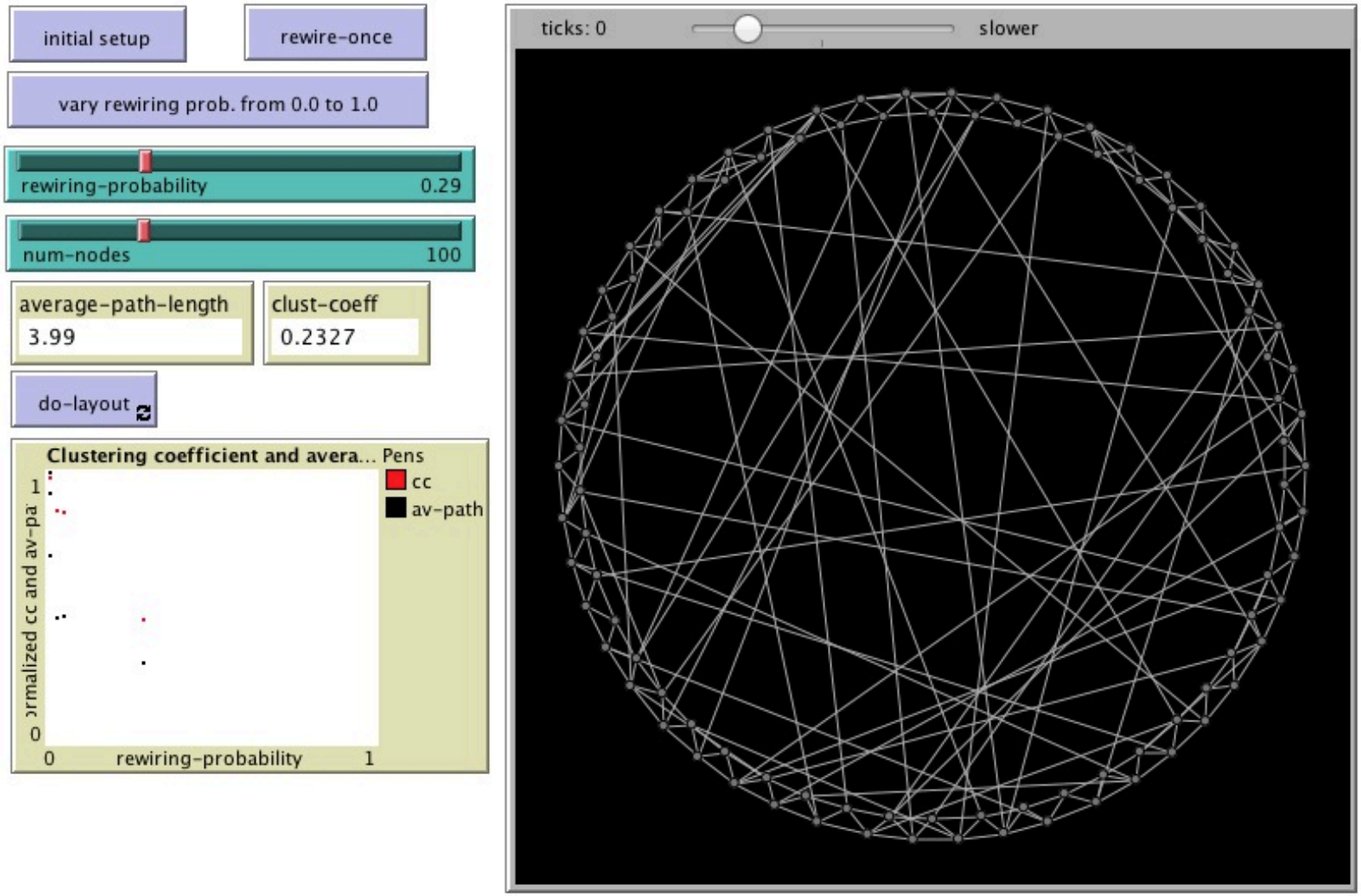
1% of links rewired

10% of links rewired

Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Trying this with NetLogo

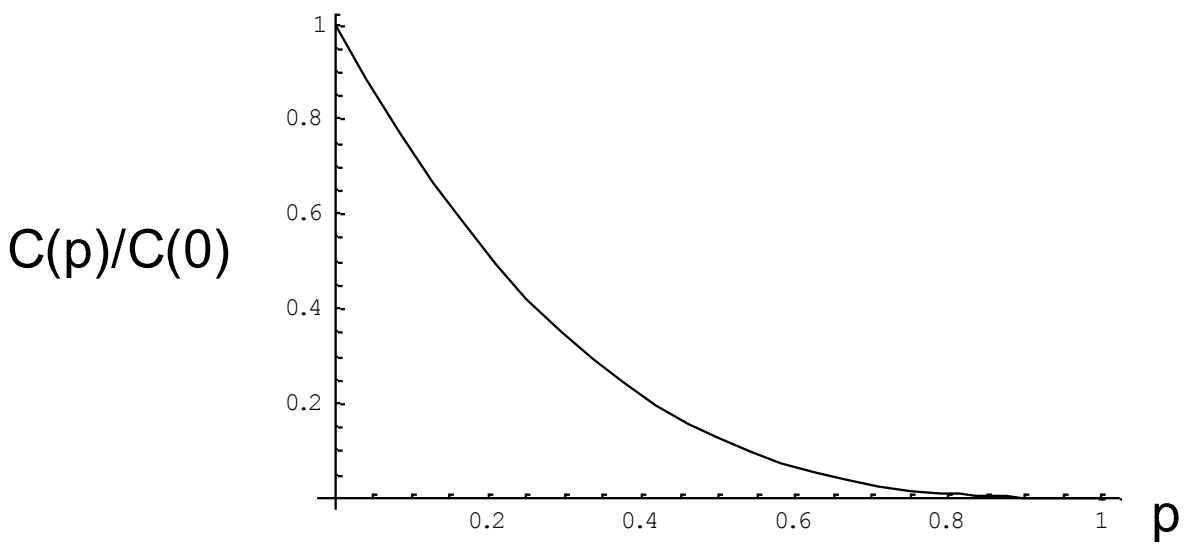
http://web.stanford.edu/class/cs224w/NetLogo/SmallWorldWS.nlogo



WS model clustering coefficient

- The probability that a connected triple stays connected after rewiring
  - probability that none of the 3 edges were rewired $(1-p)^3$
  - probability that edges were rewired back to each other very small, can ignore

- Clustering coefficient = $C(p) = C(p=0)*(1-p)^3$



Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Quiz Q

- Which of the following is a description matching a small-world network?

(a)Its average shortest path is close to that of an Erdos-Renyi graph
(b)It has many closed triads
(c)It has a high clustering coefficient
(d)It has a short average path length

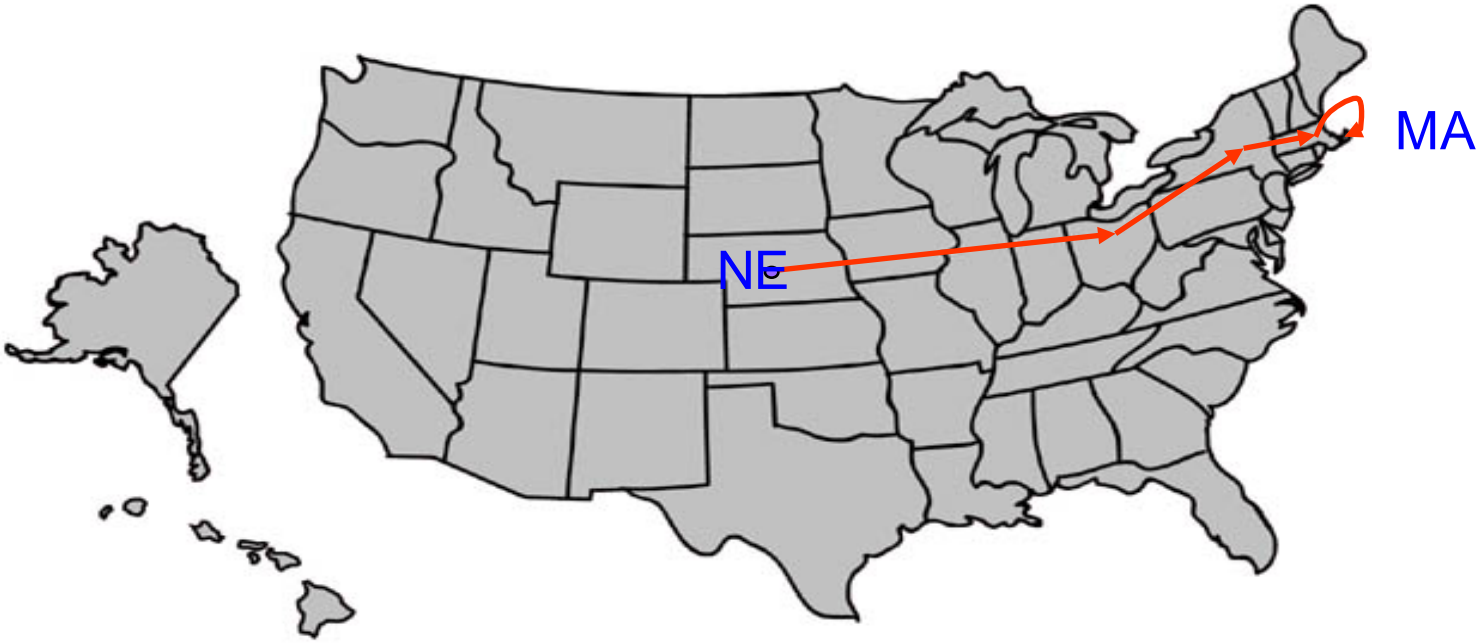
WS Model: What's missing?

- Long range links not as likely as short range ones
- Hierarchical structure / groups
- Hubs

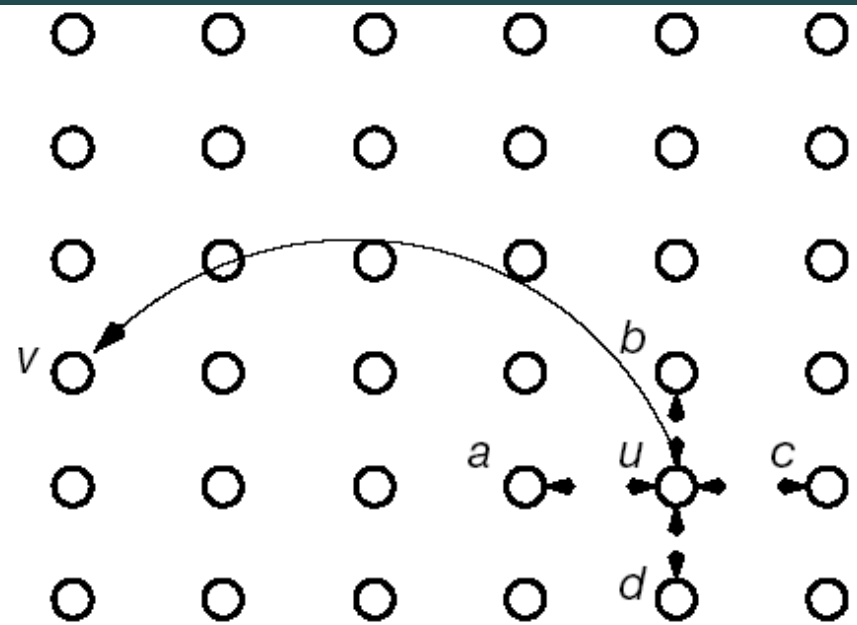
Ties and geography

"The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain"

S.Milgram 'The small world problem', Psychology Today 1,61,1967



Kleinberg's geographical small world model



nodes are placed on a lattice and connect to nearest neighbors

exponent that will determine navigability

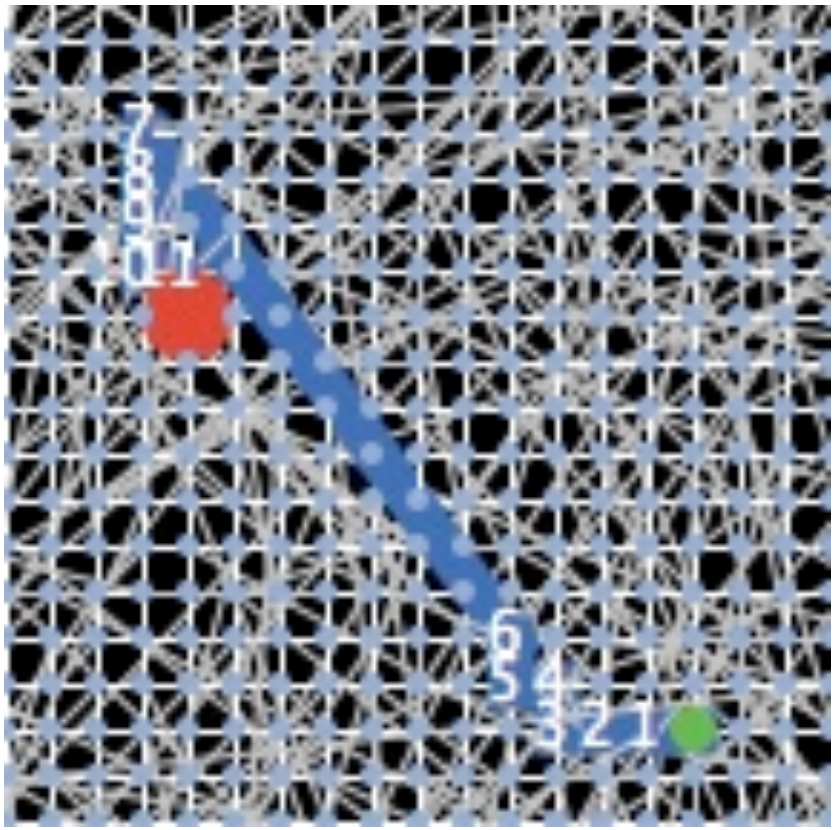
additional links placed with

$$p(\text{link between u and v}) = (\text{distance}(u,v))^{-r}$$

Source: Kleinberg, 'The Small World Phenomenon, An Algorithmic Perspective' (Nature 2000).

NetLogo demo

- how does the probability of long-range links affect search?



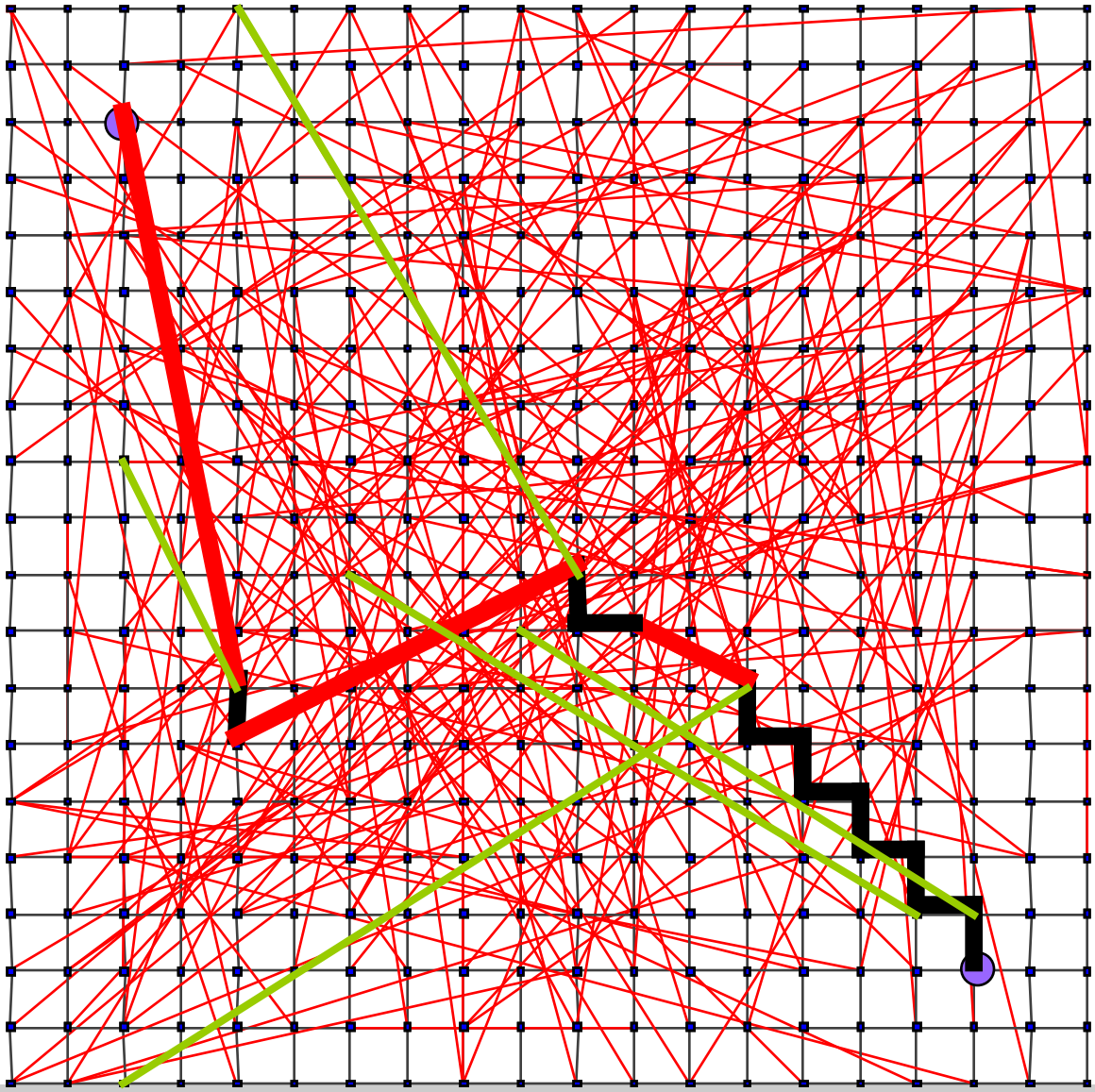
http://web.stanford.edu/class/cs224w/NetLogo/SmallWorldSearch.nlogo

geographical search when network lacks locality

When **r=0**, links are randomly distributed, ASP ~ **log(n)**, n size of grid

When **r=0**, any decentralized algorithm is at least $\mathbf{a_0 n^{2/3}}$

$p \sim p_0$

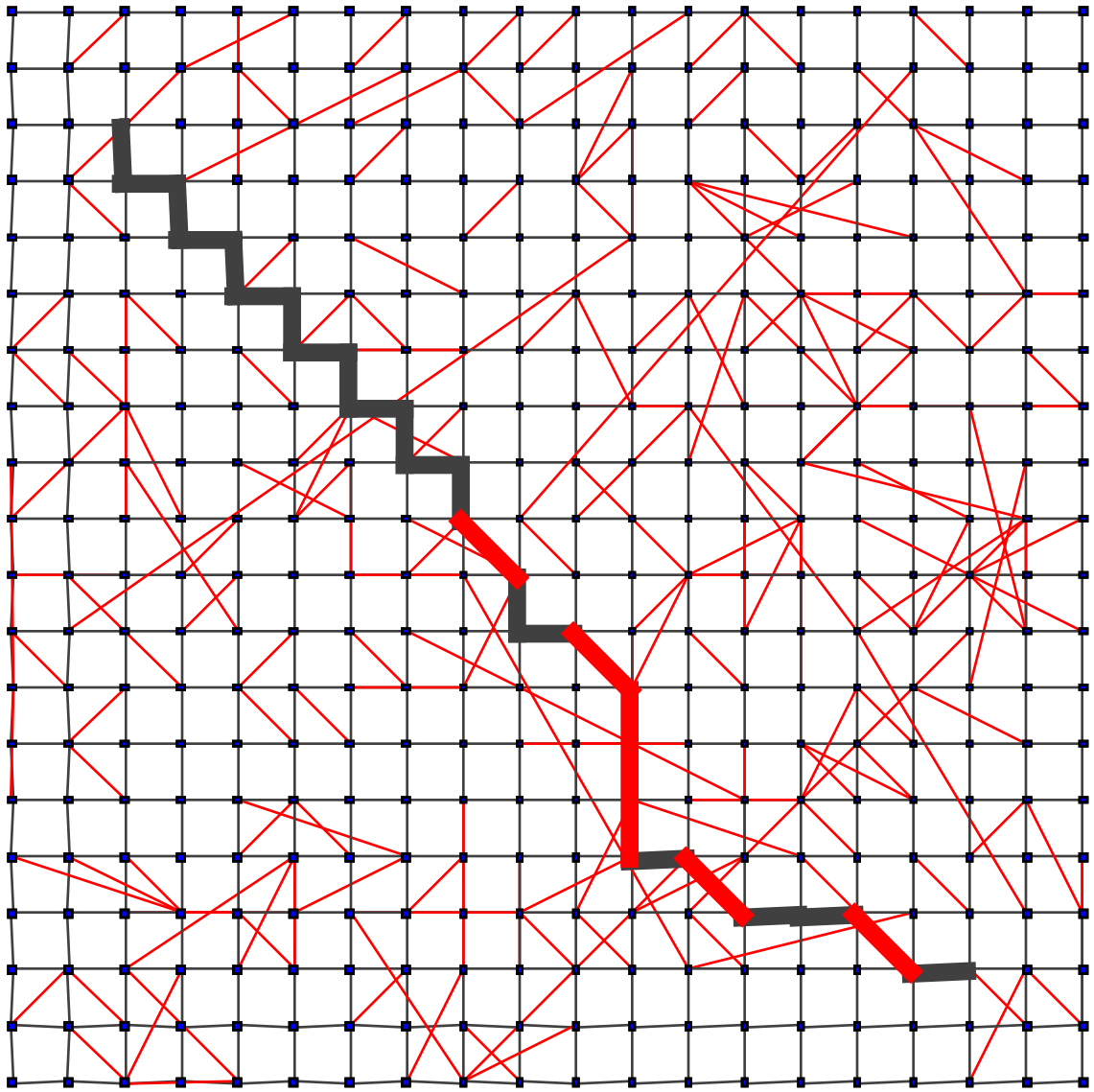


When **r<2**, expected time at least $\alpha_r n^{(2-r)/3}$

Overly localized links on a lattice

When **r>2** expected search time ~ $\mathbf{N^{(r-2)/(r-1)}}$

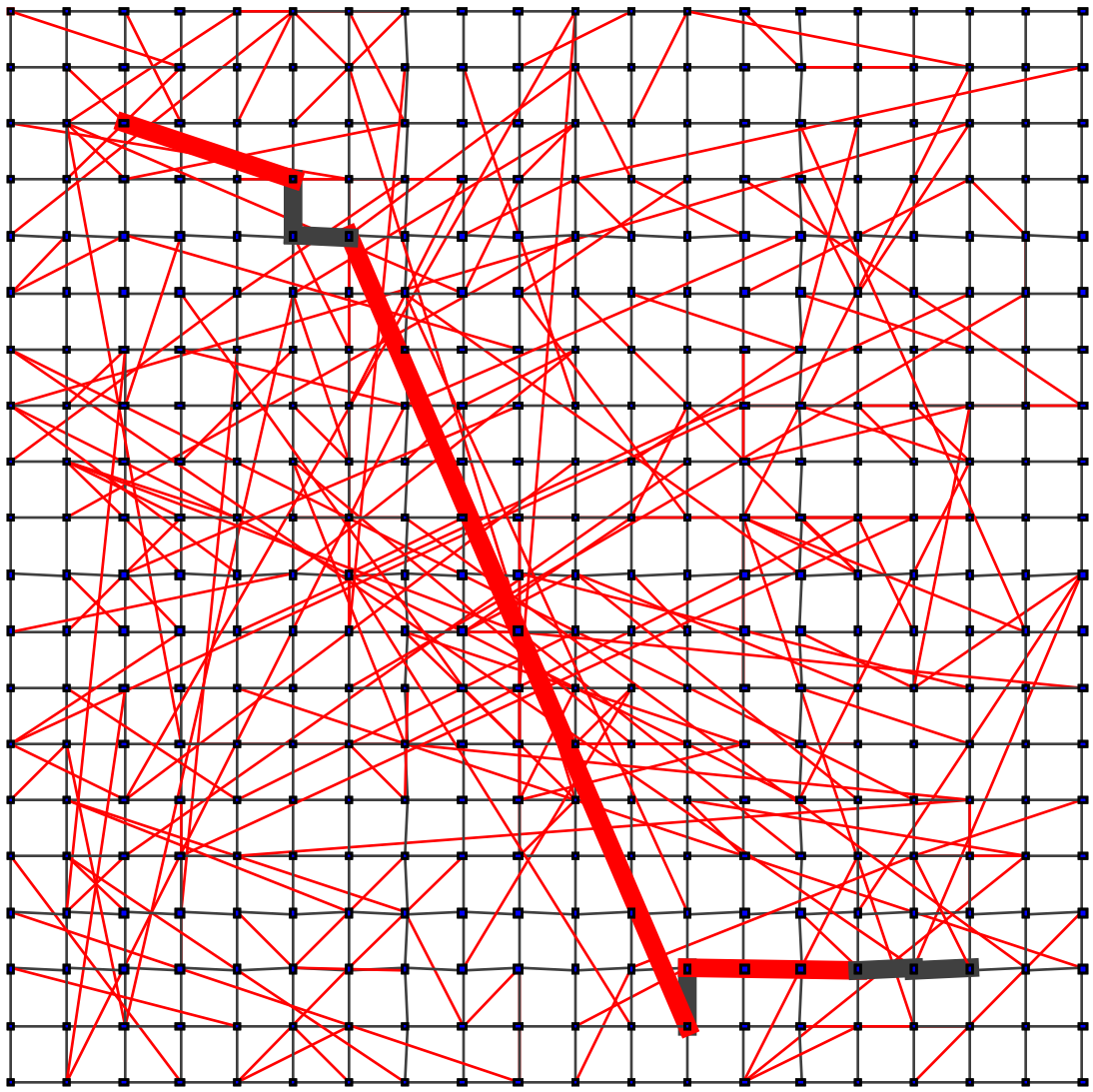
$$p \sim \frac{1}{d^4}$$



Just the right balance

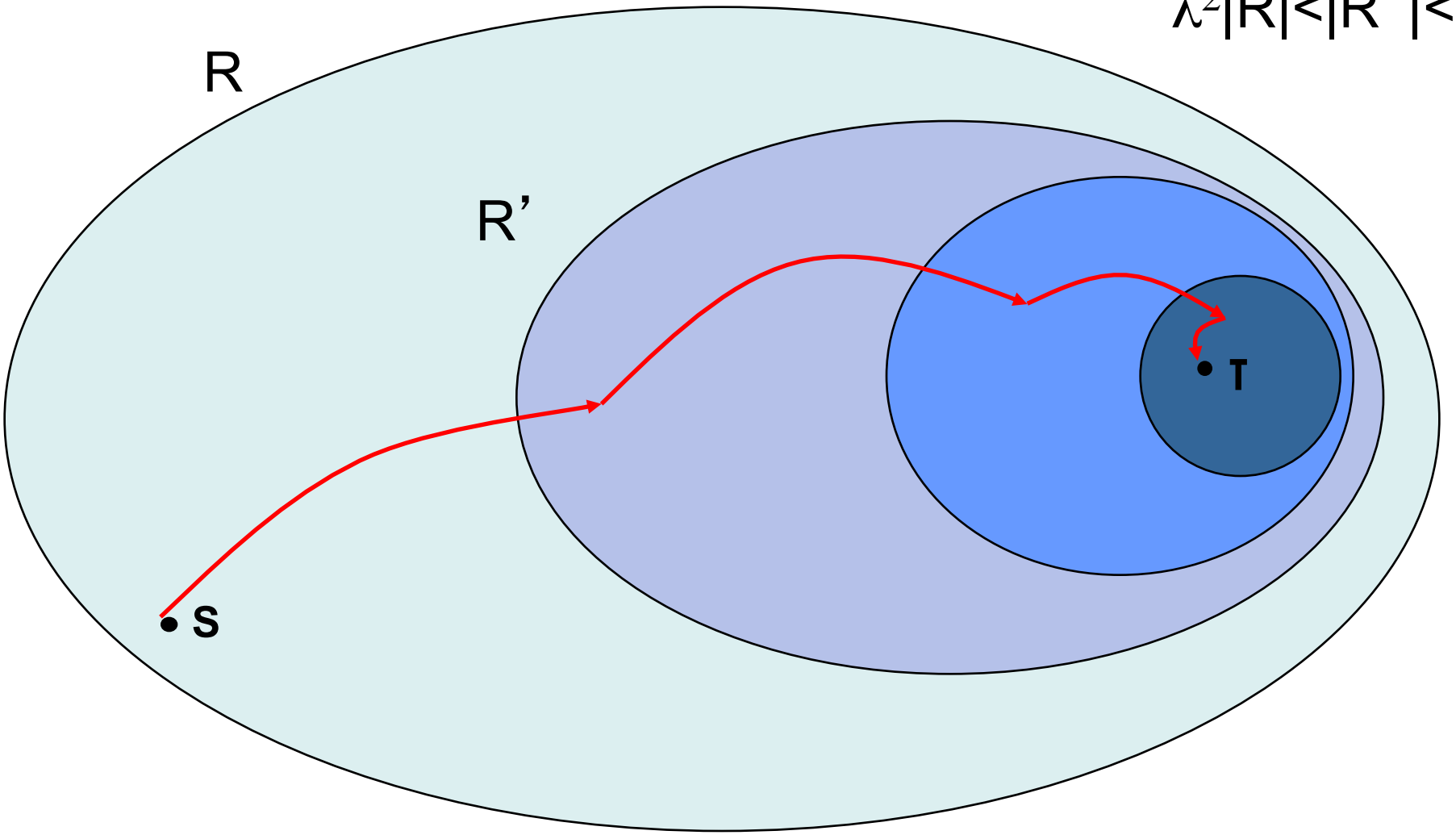
When **r=2,** expected time of a DA is at most **C (log N)^2**

$$p \sim \frac{1}{d^2}$$



Navigability

$\lambda^2|R|<|R'|<\lambda|R|$



$k = c \log^2 n$

calculate probability that s fails to have a link in R'

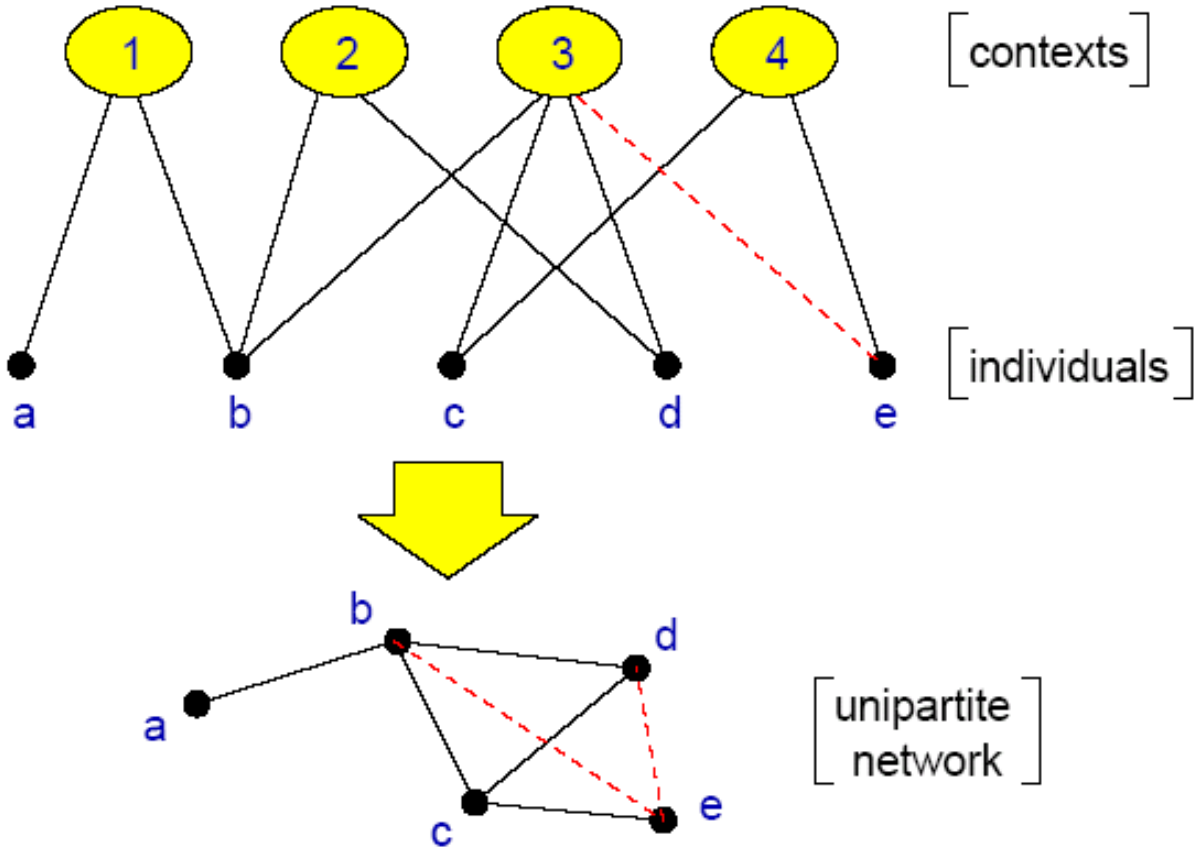
Quiz Q:

- What is true about a network where the probability of a tie falls off as distance^{-2}

(a) Large networks cannot be navigated
(b) A simple greedy strategy (pass the message to the neighbor who is closest to the target) is sufficient
(c) There are fewer long range ties than short range ones
(d) If the number of nodes doubles, the average shortest path will be twice as long

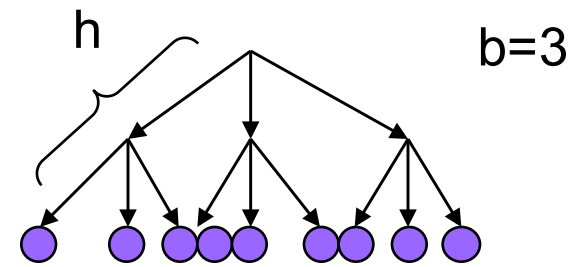
Origins of small worlds: group affiliations

Social distance—Bipartite networks:



hierarchical small-world models: Kleinberg

<u>Hierarchical network models:</u>



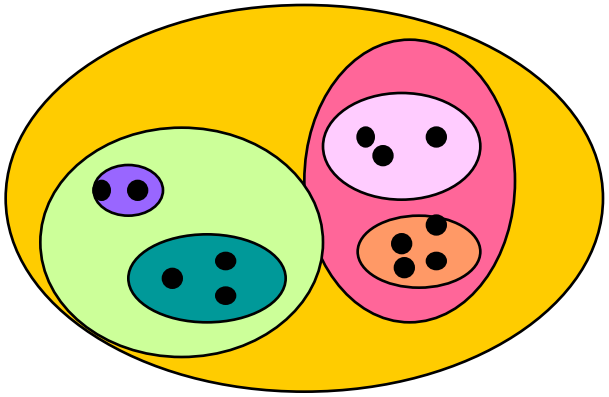
Individuals classified into a hierarchy,
h_{ij} = height of the least common ancestor.

$$p_{ij} \; : \; b^{-\alpha h_{ij}}$$

e.g. state-county-city-neighborhood
industry-corporation-division-group

<u>Group structure models:</u>
Individuals belong to nested groups
q = size of smallest group that v,w belong to

$$f(q) \sim q^{-\alpha}$$

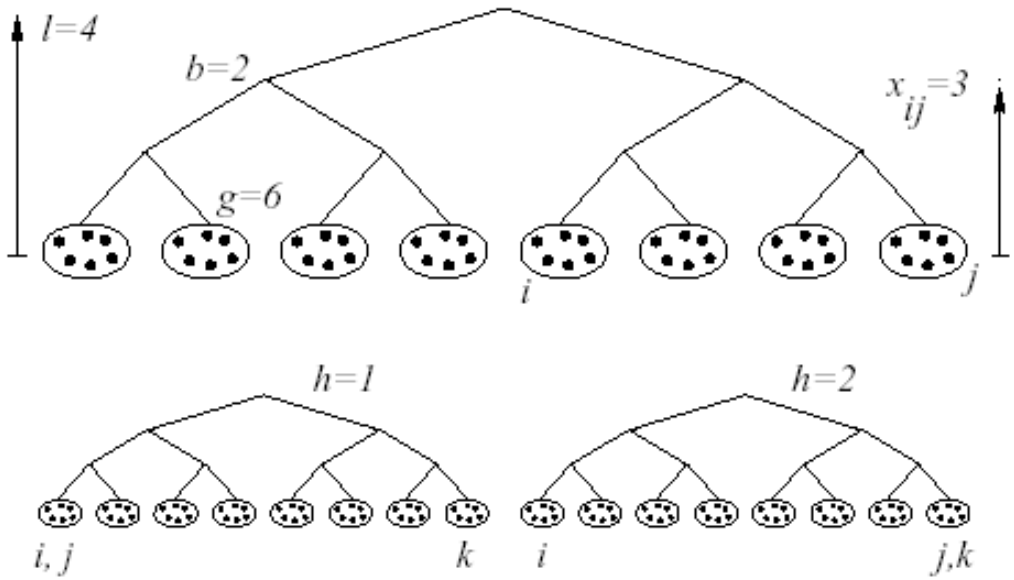


Source: <u>Kleinberg, 'Small-World Phenomena and the Dynamics of Information'</u> NIPS 14, 2001.

hierarchical small-world models: WDN

Watts, Dodds, Newman (Science, 2001)

individuals belong to hierarchically nested groups

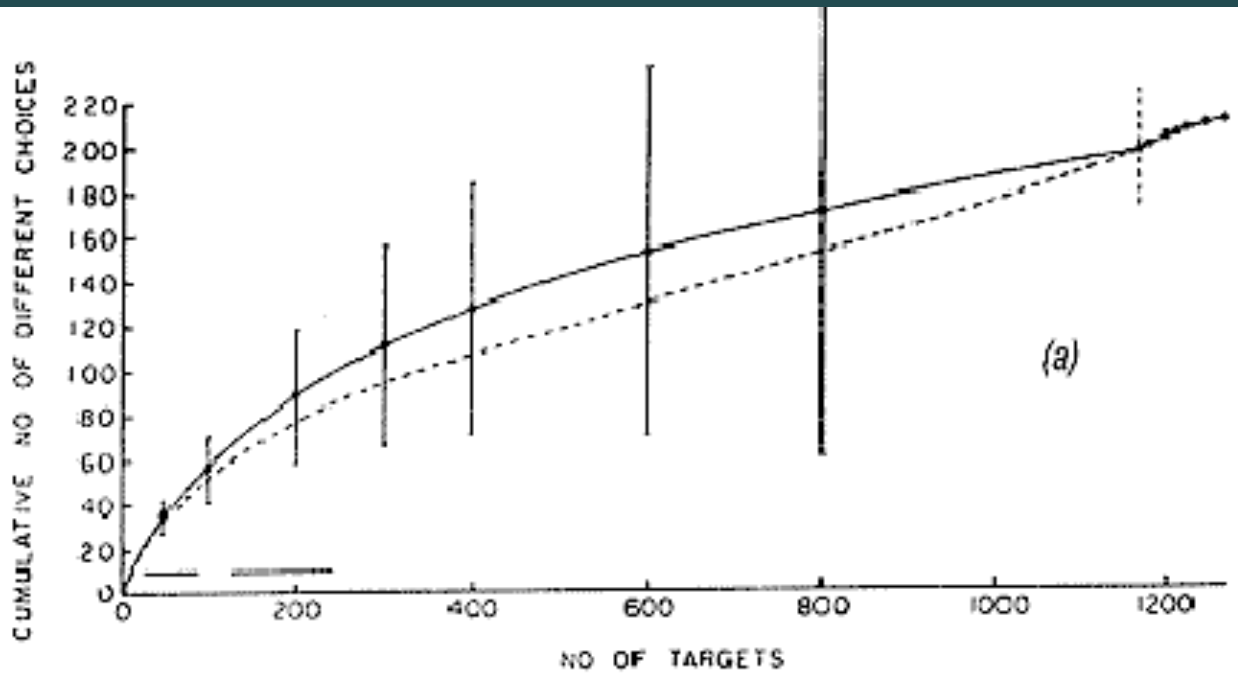


$p_{ij} \sim \exp(-\alpha\, x)$

multiple independent hierarchies h=1,2,..,H coexist corresponding to occupation, geography, hobbies, religion…

Source: Identity and Search in Social Networks: Duncan J. Watts, Peter Sheridan Dodds, and M. E. J. Newman; Science 17 May 2002 296: 1302-1305. < http://arxiv.org/abs/cond-mat/0205383v1 >

Navigability and search strategy: Reverse small world experiment



- Killworth & Bernard (1978):
- Given hypothetical targets (name, occupation, location, hobbies, religion…) participants choose an acquaintance for each target
- based on (most often) occupation, geography
- only 7% because they "know a lot of people"
- Simple greedy algorithm: most similar acquaintance
- two-step strategy rare

Source: 1978 Peter D. Killworth and H. Russell Bernard. The Reverse Small World Experiment Social Networks 1:159–92.

Navigability and search strategy: Small world experiment @ Columbia

Successful chains disproportionately used

- weak ties (Granovetter)
- professional ties (34% vs. 13%)
- ties originating at work/college
- target's work (65% vs. 40%)

. . . and disproportionately avoided

- hubs (8% vs. 1%) (+ no evidence of funnels)
- family/friendship ties (60% vs. 83%)

Strategy: Geography -> Work

Search in power-law networks

Motivation

Power-law (PL) networks, social and P2P

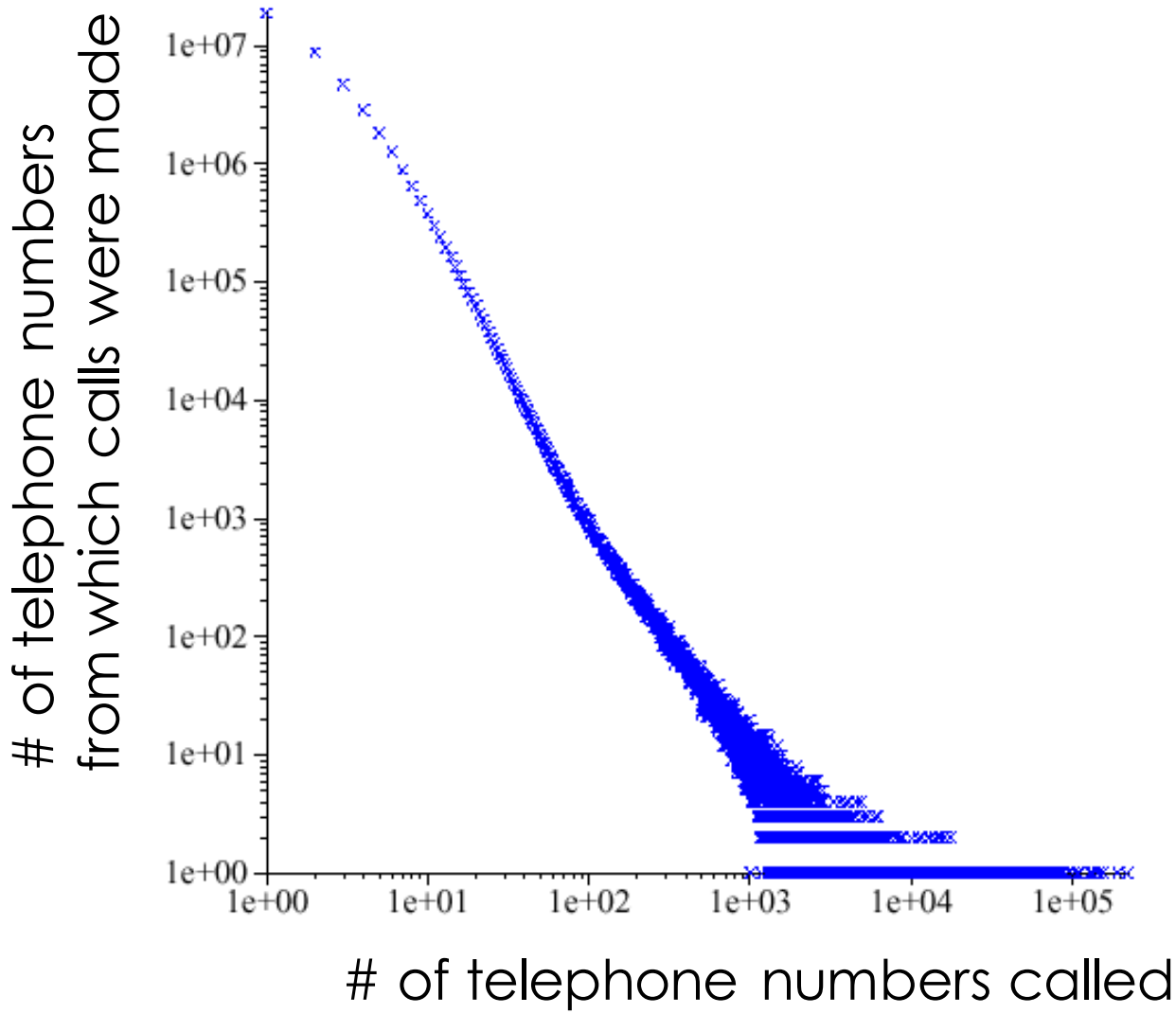
Analysis of scaling of search strategies in PL networks

Simulation

artificial power-law topologies, real Gnutella networks

How do we search?

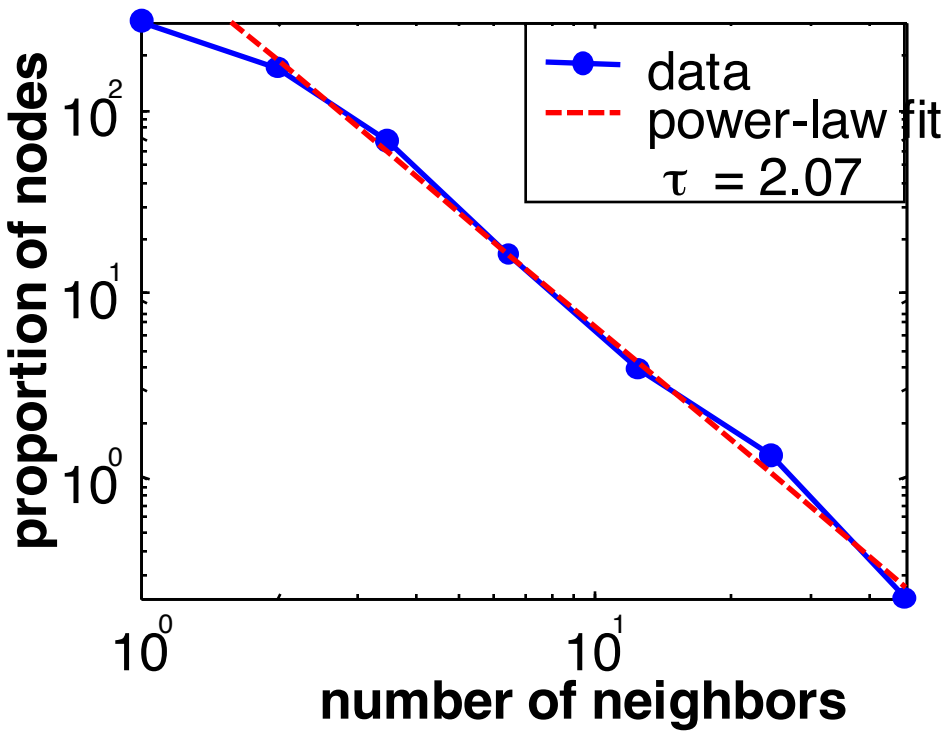
AT&T Call Graph



Aiello et al. STOC '00

Gnutella network

power-law link distribution



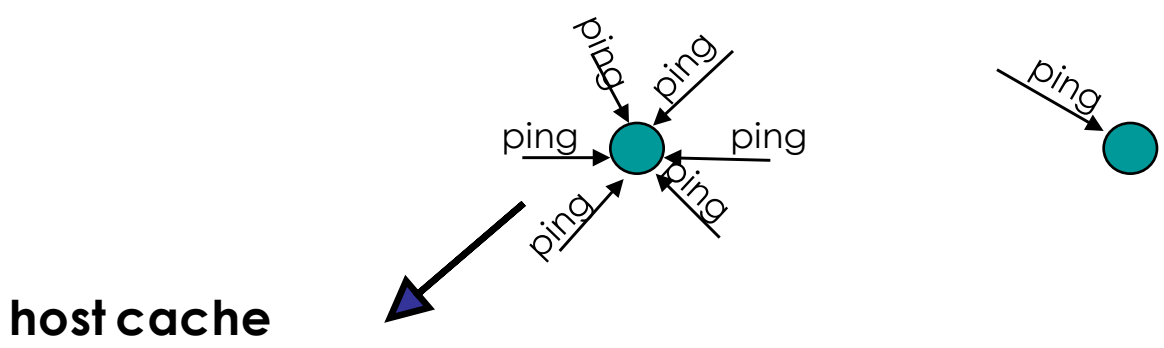
summer 2000,
data provided by Clip2

Preferential attachment model

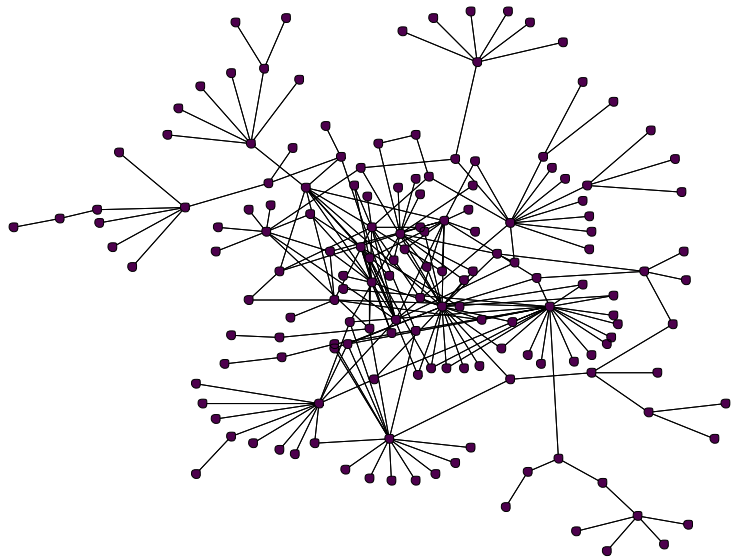
Nodes join at different times

The more connections a node has, the more likely it is to acquire new connections

Growth process produces power-law network



Gnutella and the bandwidth barrier

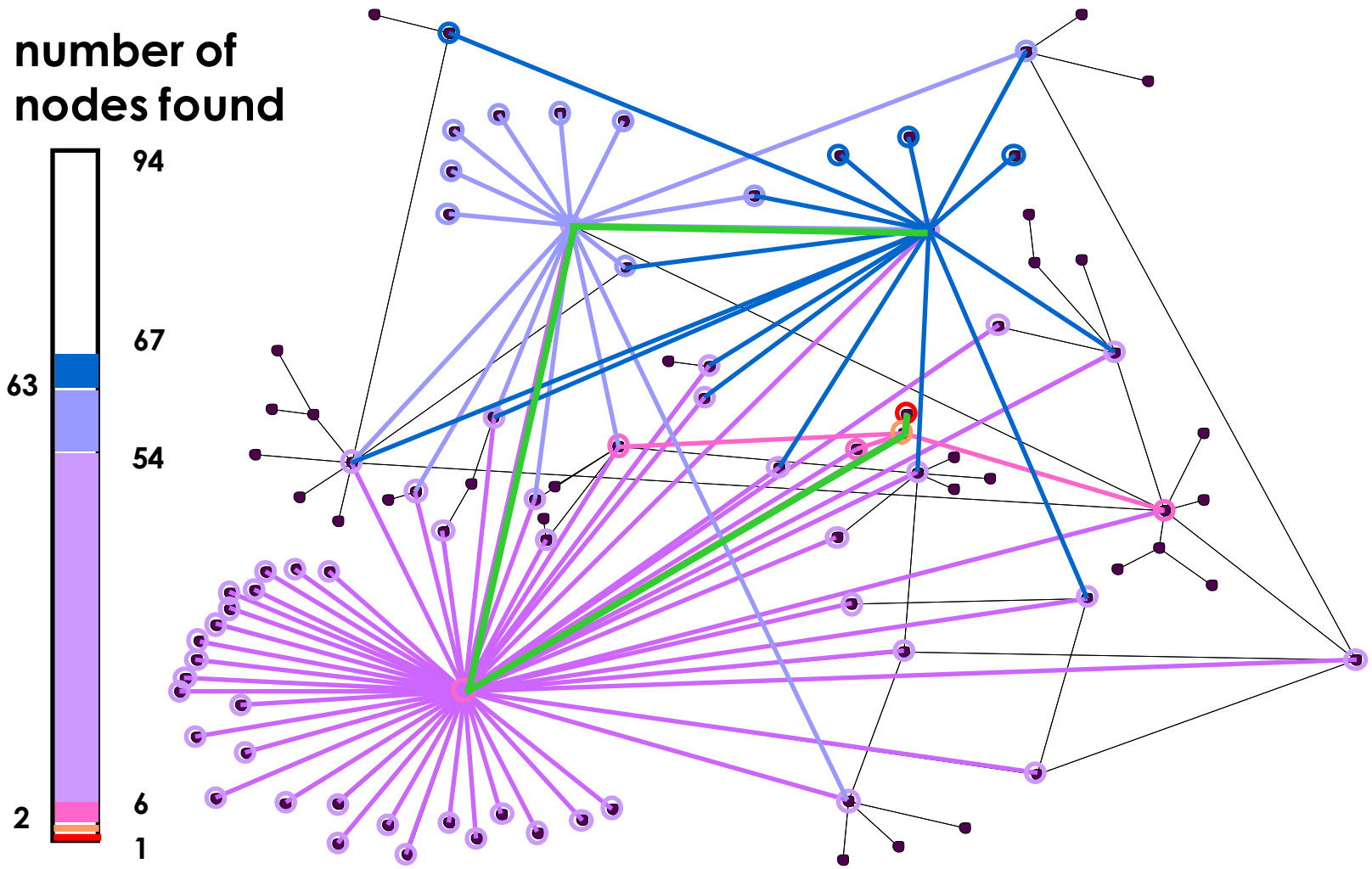


file sharing w/o a central index

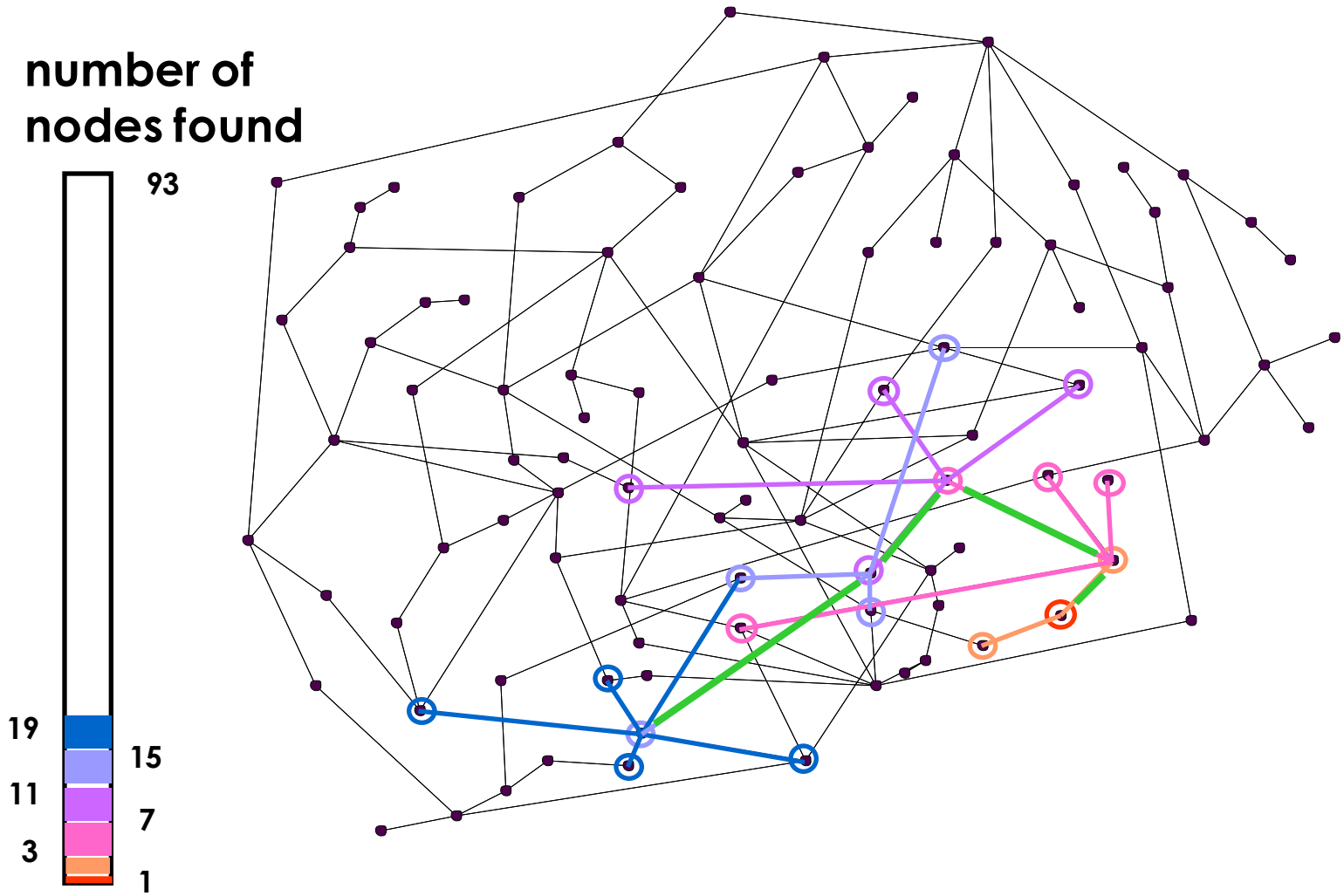
queries broadcast to every node within radius *ttl*
⇒ as network grows, encounter a bandwidth barrier (dial up modems cannot keep up with query traffic, fragmenting the network)

Clip 2 report
Gnutella: To the Bandwidth Barrier and Beyond
http://www.clip2.com/gnutella.html#q17

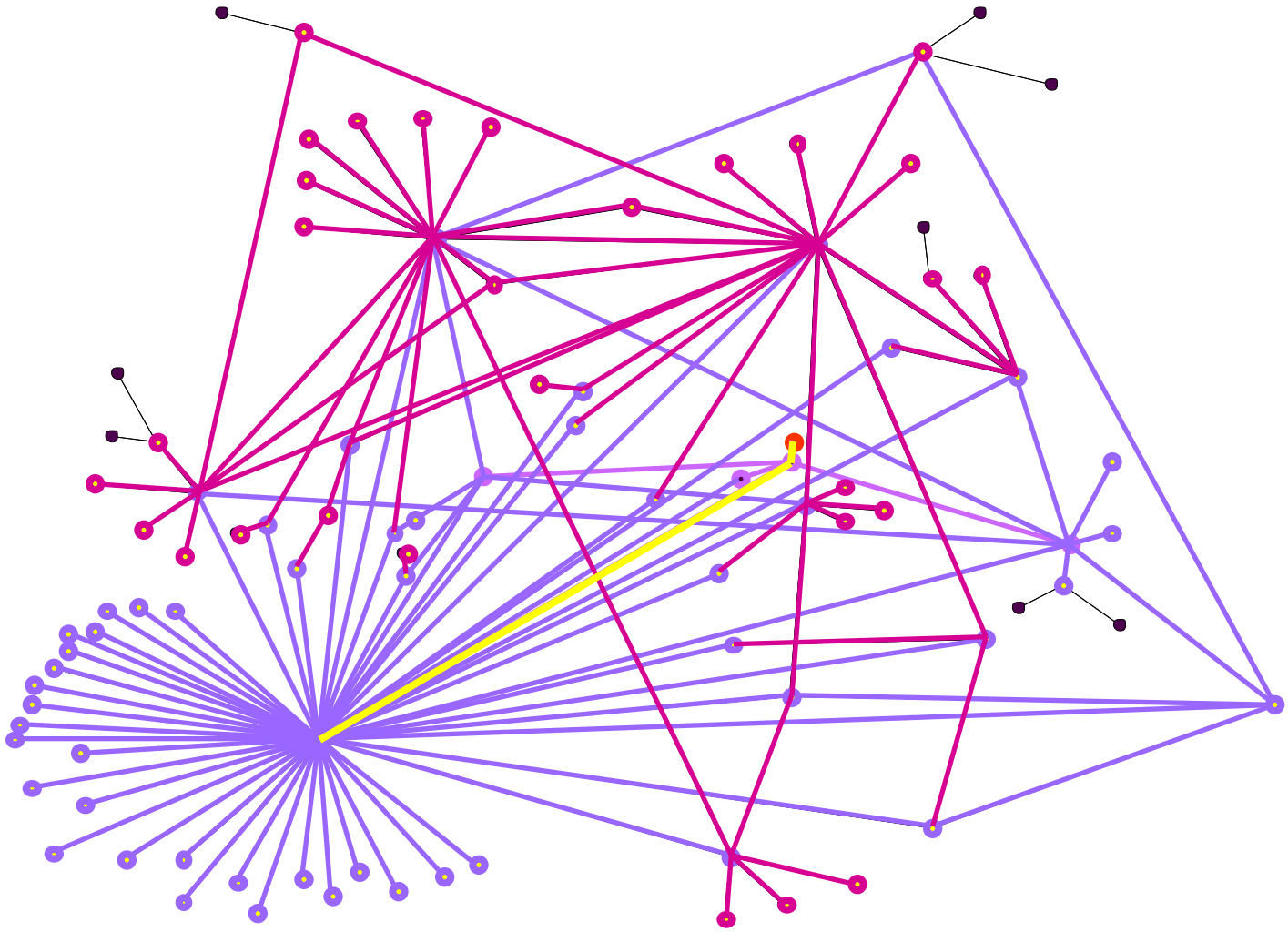
power-law graph



Poisson graph



Search with knowledge of 2nd neighbors



Outline of search strategy

pass query onto only **one** neighbor at each step

OPTIONS

requires that nodes sign query
- avoid passing message onto a node twice

requires knowledge of one's neighbors degree
- pass to the highest degree node

requires knowledge of one's neighbors neighbors
- route to 2nd degree neighbors

Generating functions

- M.E.J. Newman, S.H. Strogatz, and D.J. Watts
- 'Random graphs with arbitrary degree distributions and their applications', PRE, cond-mat/0007235

- Generating functions for degree distributions

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$$

- Useful for computing moments of degree distribution,
- component sizes, and average path lengths

Introducing cutoffs

$k_{\max} < N - 1$ a node cannot have more connections than there are other nodes

This is important for exponents close to 2

$$\sum_{1}^{\infty} p_k = \sum_{1}^{\infty} C_\tau \frac{1}{x^\tau} = 1 \qquad C_2 = \frac{6}{\pi^2}$$

$$p(k > 1000, \tau = 2) = \sum_{1000}^{\infty} p_k \sim 0.001$$

Probability that none of the nodes in a 1,000 node graph has 1000 or more neighbors:

$$(1 - p(k > 1000, \tau = 2))^{1000} \sim 0.36$$

without a cutoff, for $\tau = 2$
have > 50% chance of observing a node with more neighbors than there are nodes

for $\tau = 2.1$, have a 25% chance

Selecting from a variety of cutoffs

1. $k_{\max} < N$

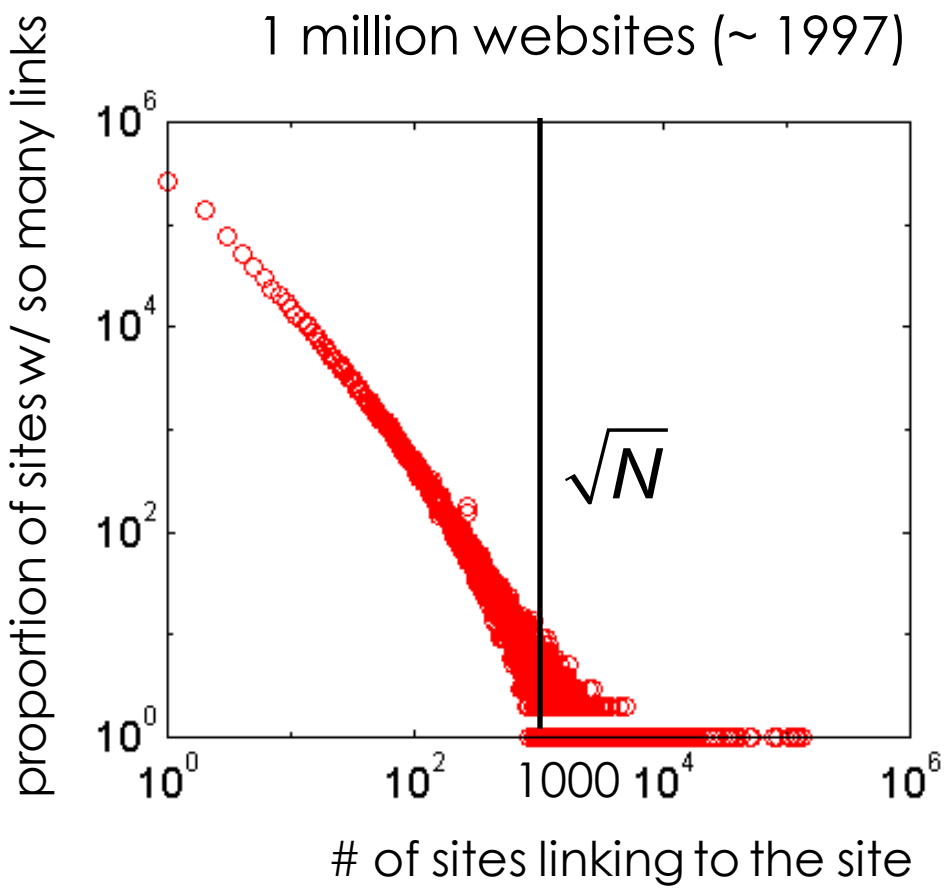
2. $p_k = Ck^{-\tau} e^{-k/\kappa}$ Newman et al.

3. $p_k = \begin{cases} Ck^{-\tau} & k < (CN)^{1/\tau} \\ 0 & \text{otherwise} \end{cases}$

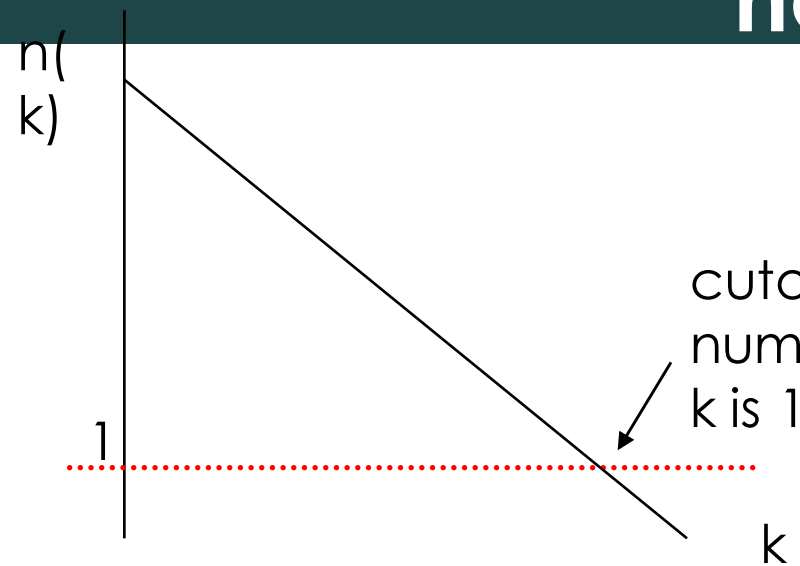
Aiello et al.

Generating Function

$$G_0(x) = C \sum_{k=1}^{(CN)^{1/\tau}} k^{-\tau} x^k$$



Aiello's 'conservative' vs. Havlin's 'natural' cutoff

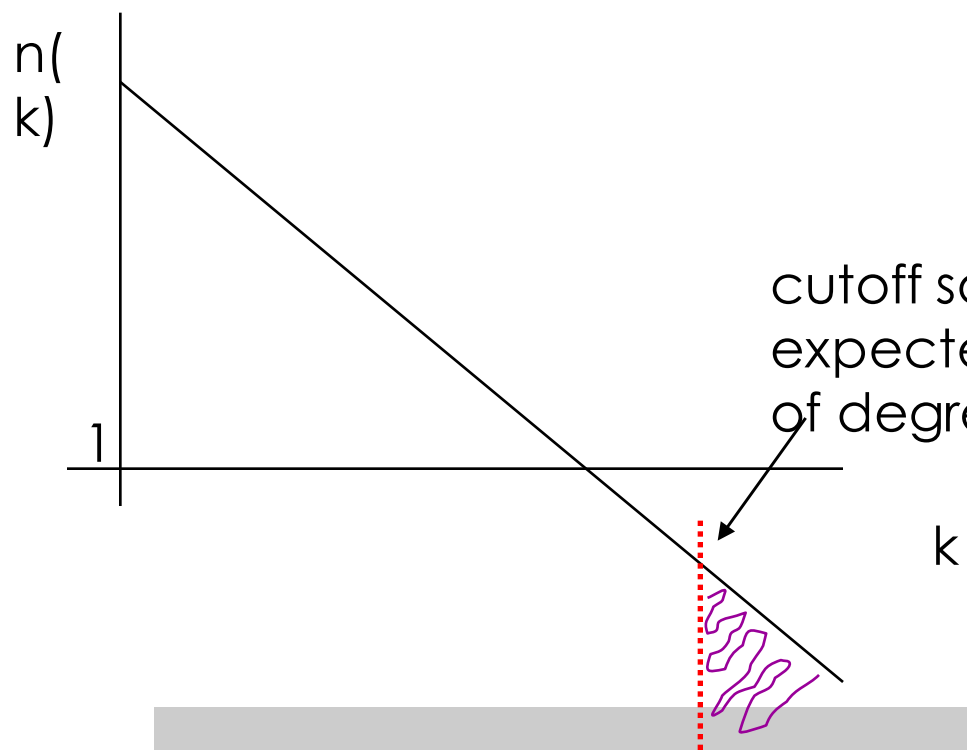


cutoff where expected number of nodes of degree k is 1

$$N * p_k = 1$$

$$Ck^{-\tau} = N^{-1}$$

$$k \sim N^{\frac{1}{\tau}}$$



cutoff so that expected number of nodes of degree > k is 1

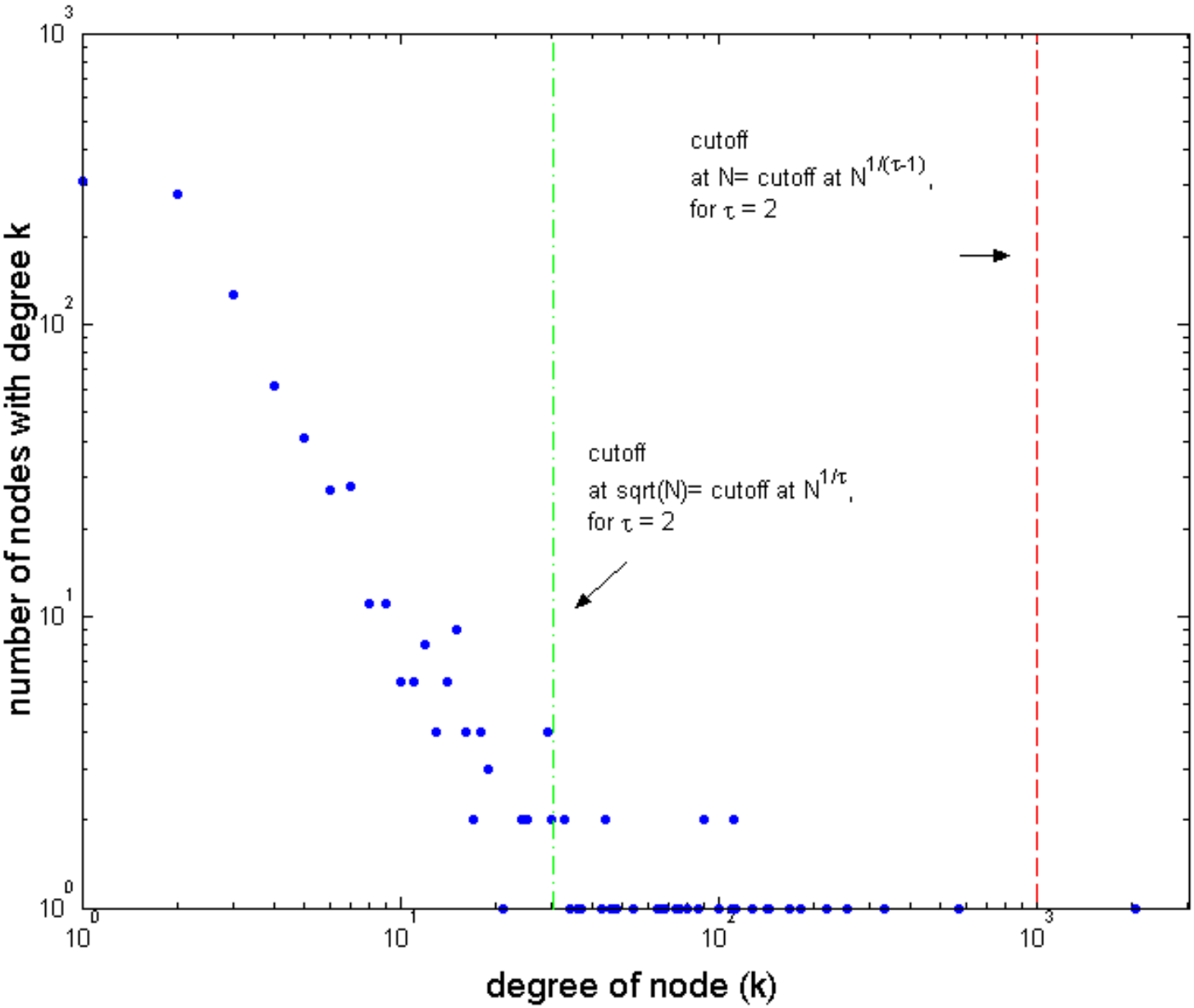
$$N * \sum_{k=k_{\max}}^{\infty} p_k = 1$$

$$\int_{k=k_{\max}}^{\infty} ck^{-\tau} \sim N^{-1}$$

$$k_{\max}^{1-\tau} \sim N^{-1}$$

$$k_{\max} \sim N^{\frac{1}{\tau-1}}$$

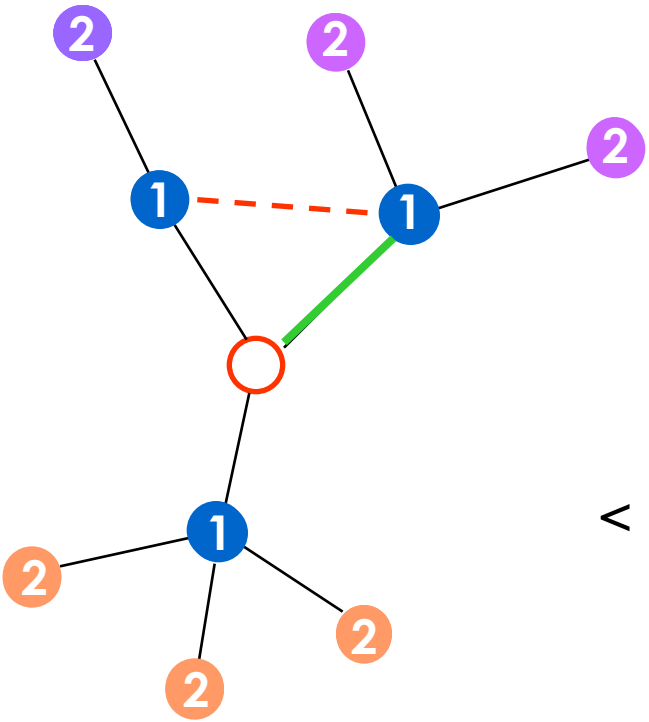
The imposed cutoff can have a dramatic effect on the properties of the graph



Generating functions for degree distributions

Random graphs with arbitrary degree distributions and their applications

by Newman, Strogatz & Watts



$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$$ is a generating function

$$p_k \sim k^{-\tau}$$ is the probability that a randomly chosen vertex has degree k

$$< k >= \sum_k k p_k = G_0^{'}(1)$$ is the expected degree of a randomly chosen vertex

$$G_1(x) = \frac{G_0^{'}(x)}{G_0^{'}(1)}$$ is the distribution of remaining outgoing edges following and edge

$$z_2 = G_0^{'}(1) G_1^{'}(1)$$ is the expected number of second degree neighbors

assuming neighbors don't share edges

search with knowledge of first neighbors

$$G_0(x) = c\sum_{1}^{k_{\max}} k^{-\tau} x^k$$

Generating function with cutoff

$$G_0'(x) = \frac{\partial}{\partial x} G_0(x) = c\sum_{1}^{k_{\max}} k^{1-\tau} x^{k-1}$$

Average degree of vertex

$$G_0'(1) = \langle k \rangle = c\sum_{1}^{k_{\max}} k^{1-\tau} \; : \; \int_{1}^{k_{\max}} k^{1-\tau} dk = \frac{1}{\tau - 2}\left(1 - k_{\max}^{2-\tau}\right)$$

$$G_1'(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{c}{G_0'(1)} \frac{\partial}{\partial x} \sum_{1}^{k_{\max}} k^{1-\tau} x^{k-1}$$

Average number of neighbors following an edge

$$= \frac{c}{G_0'(1)} \sum_{2}^{k_{\max}} k^{1-\tau}(k-1)x^{k-2}$$

for $2<\tau<3$, and $k_{max} \sim N^{a}$, decreases with N

constant in N

$$G_1'(1) = \frac{1}{G_0'(1)} \frac{k_{\max}^{3-\tau}(\tau - 2) - 2^{2-\tau}(\tau - 1) + k_{\max}^{2-\tau}(3 - \tau)}{(\tau - 2)(3 - \tau)}$$

search with knowledge of first neighbors (cont'd)

$$z_{1B} = G_1^{'}(1) : \frac{1}{G_0^{'}(1)} \frac{k_{\max}^{3-\tau}}{(3-\tau)} = \frac{\tau - 2}{1 - k_{\max}^{2-\tau}} \frac{k_{\max}^{3-\tau}}{(3-\tau)} : k_{\max}^{3-\tau}$$

In the limit τ->2, $G_1^{'}(1) : \frac{k_{\max}}{\log(k_{\max})}$

Let's for the moment ignore the fact that as we do a random walk, we encounter neighbors
that we've seen before

$$s = \text{number of steps} = \frac{N}{z_{1B}}$$

Search time with different cutoffs

If $k_{max} = N$, $s(\tau): \frac{N}{k_{max}^{3-\tau}} = \frac{N}{N^{3-\tau}} = N^{\tau-2}, 2 < \tau < 3$

$s(2.1): N^{0.1}$

$s: \frac{N\log(k_{max})}{k_{max}} = \log(N), \tau = 2$

If $k_{max} = N^{1/(\tau-1)}$, $s(\tau): \frac{N}{k_{max}^{3-\tau}} = \frac{N}{N^{\frac{3-\tau}{\tau-1}}} = N^{2\frac{\tau-2}{\tau-1}}, 2 < \tau < 3$

$s(2.1): N^{0.18}$

$s(2): \frac{N\log(k_{max})}{k_{max}} = \log(N)$

search with knowledge of first neighbors (cont'd)

If $k_{max} = N^{1/\tau}$, s : $\frac{N}{k_{max}^{3-\tau}} = \frac{N}{(N^{\frac{1}{\tau}})^{3-\tau}} = N^{2-3/\tau}, 2 < \tau < 3$

So the best we can do is $\sqrt{N}$ for exponents close to 2

2nd neighbor random walk, ignoring overlap:

$$z_{2B} = \left[\frac{\partial}{\partial x} G_1(G_1(x))\right]_{x=1} = \left[G_1'(1)\right]^2 = \left[\frac{\tau-2}{1-k_{max}^{2-\tau}} \frac{k_{max}^{3-\tau}}{(3-\tau)}\right]^2$$

$$n_s = z_{2B}(N)$$

$$S \sim \frac{N}{z_{2B}(N)}$$

$$S(N,\tau) \sim N^{3(1-2/\tau)} \qquad S(N,\tau = 2.1) \sim N^{0.15}$$

Following the degree sequence

Go to highest degree node, then next highest, … etc.

$$z_{1D} = \int_{k_{\max}-a}^{k_{\max}} Nk^{1-\tau} dk \sim Nak_{\max}^{1-\tau}$$

a ~ s = # of steps taken

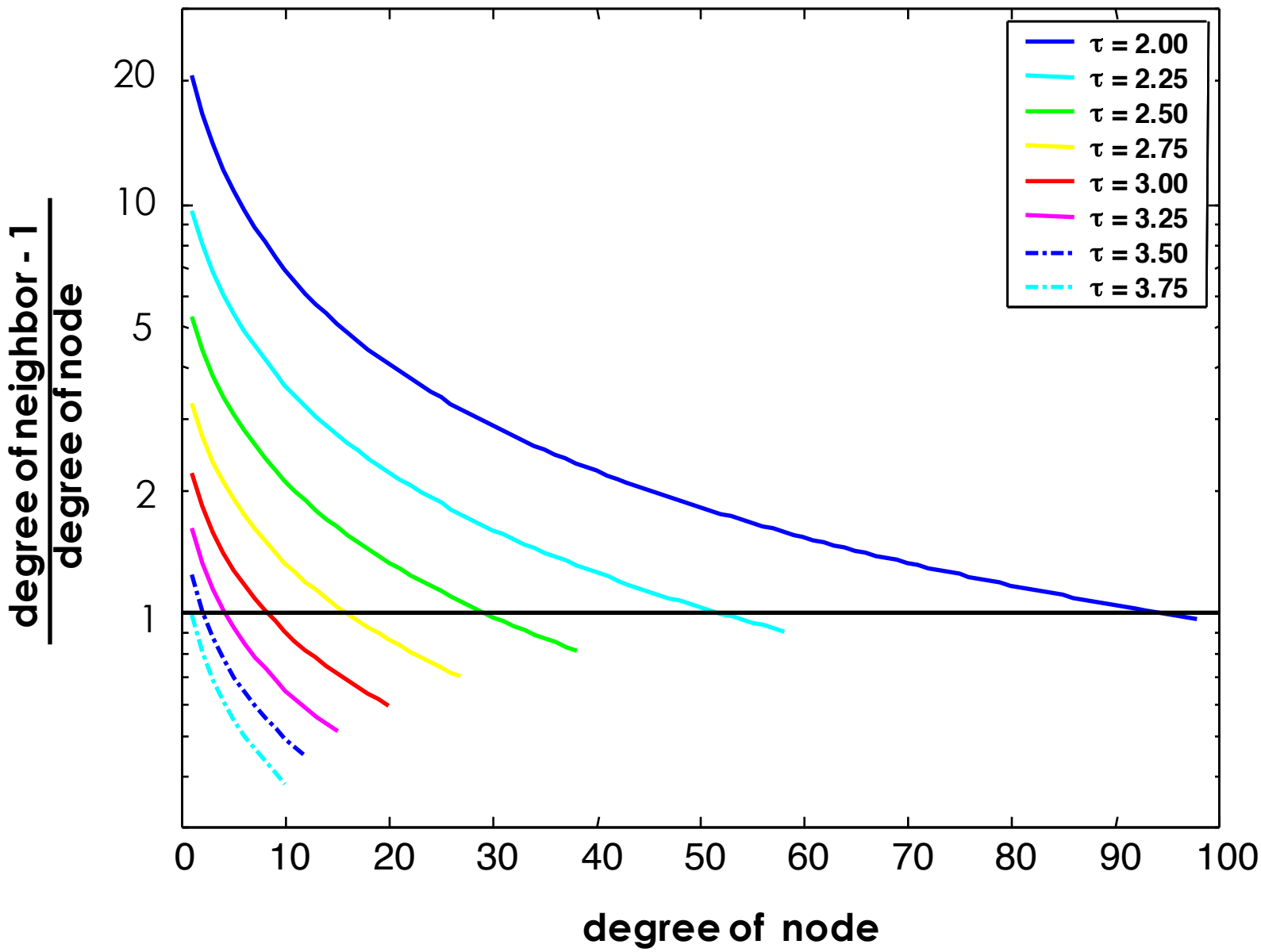
2nd neighbors, ignoring overlap:

$$z_{1D} G_1'(x) \sim Nak_{\max}^{2(2-\tau)}$$

$$s \sim k_{\max}^{2(\tau-2)} \sim N^{2-4/\tau}$$

$$S_{\deg}(N, \tau = 2.1) = N^{0.1}$$

Ratio of the degree of a node to the expected degree of its highest degree neighbor for 10,000 node power-law graphs of varying exponents



Exponents τ close to 2 required to search effectively

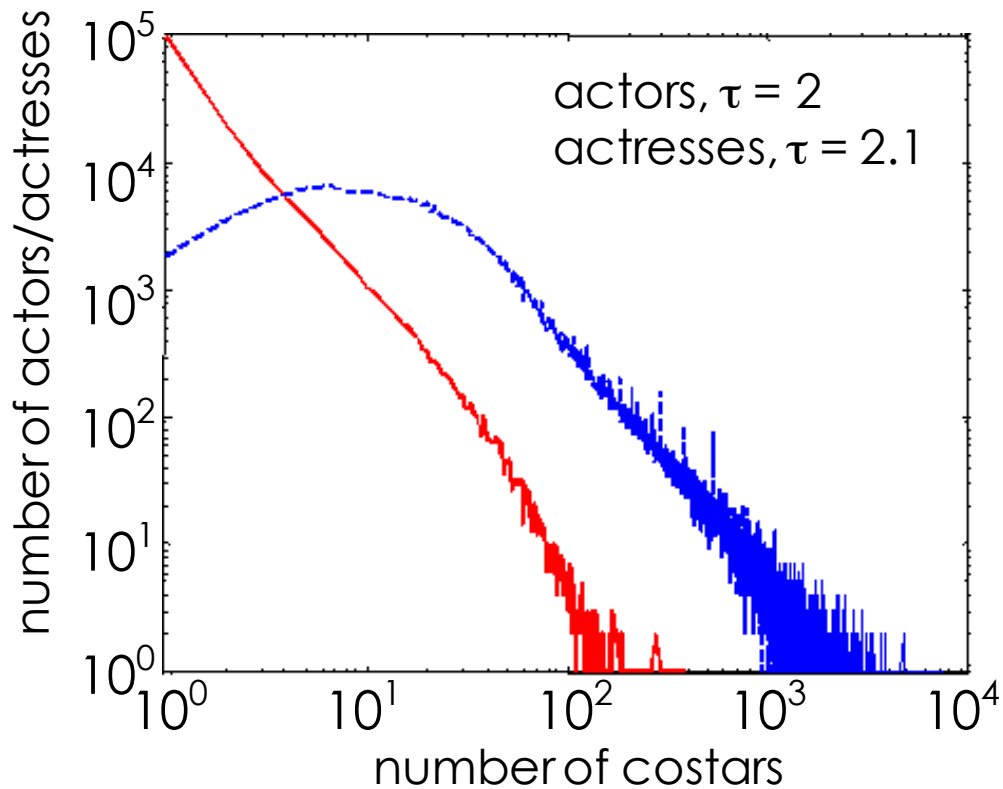
Gnutella

World Wide Web, $\tau \sim 2.0$-2.3,
high degree nodes: directories, search engine

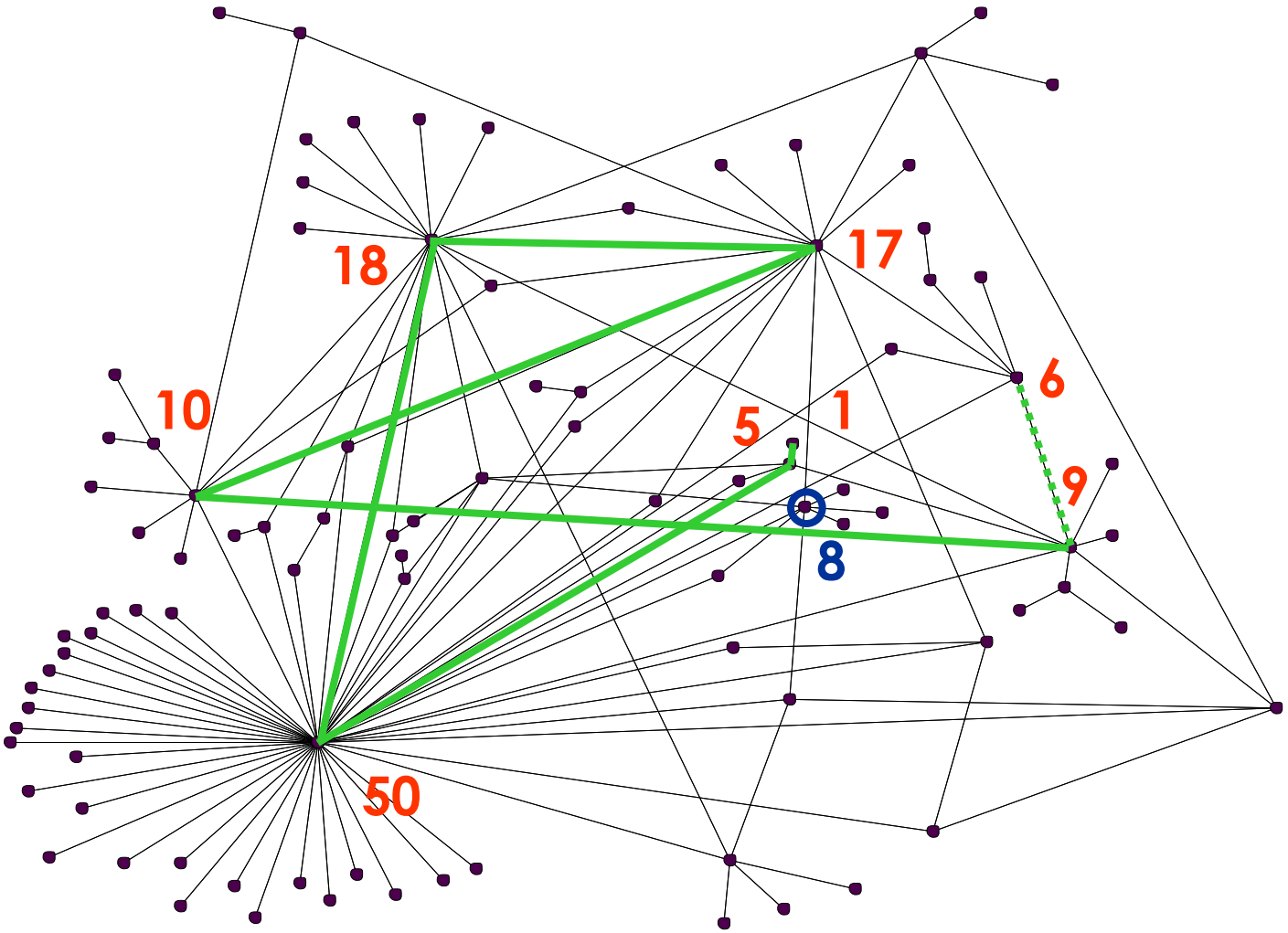
Social networks, AT&T call graph $\tau \sim 2.1$

Actor collaboration graph
(imdb database)

$\tau \sim 2.0$-2.2



Following the degree sequence

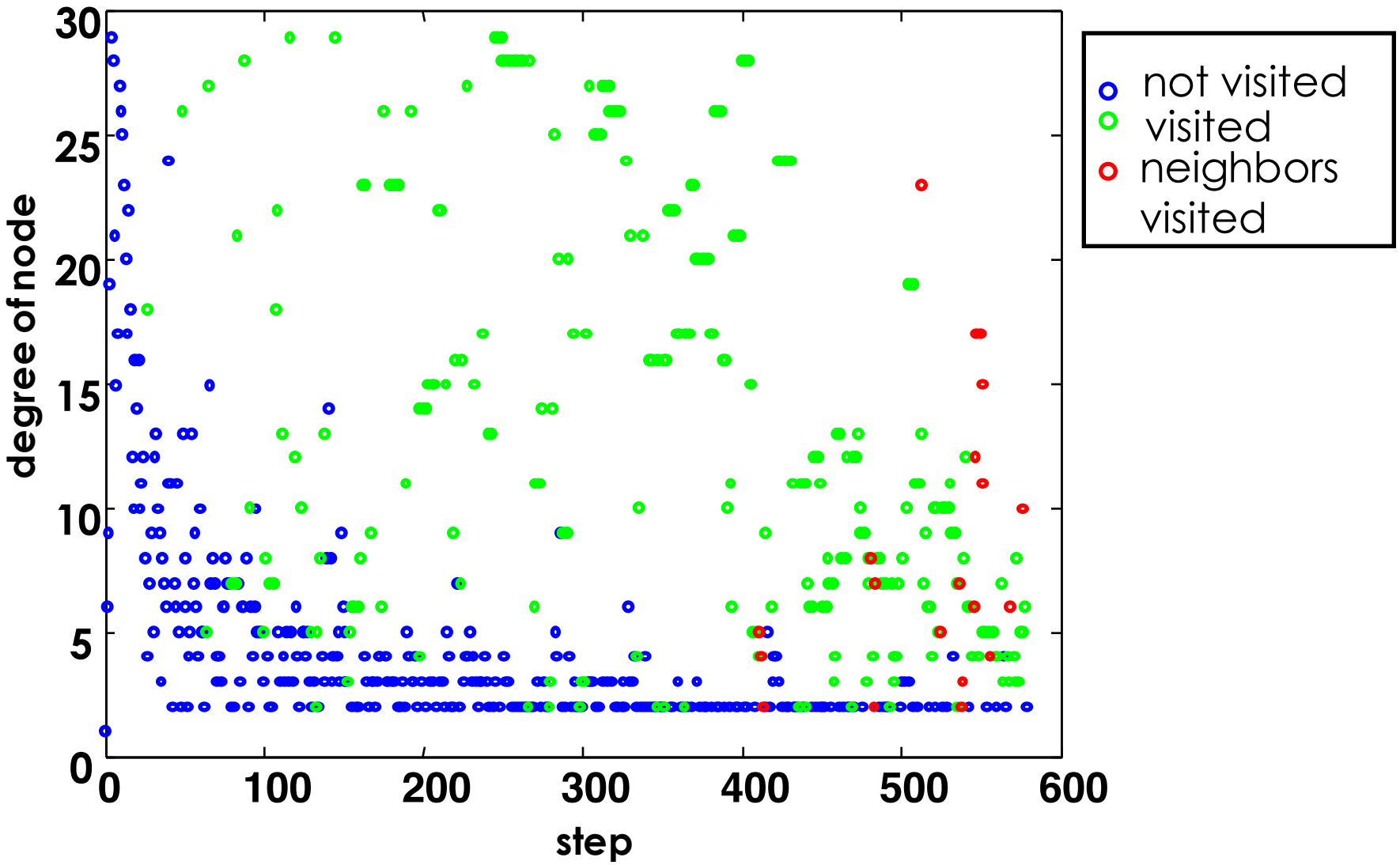


Complications

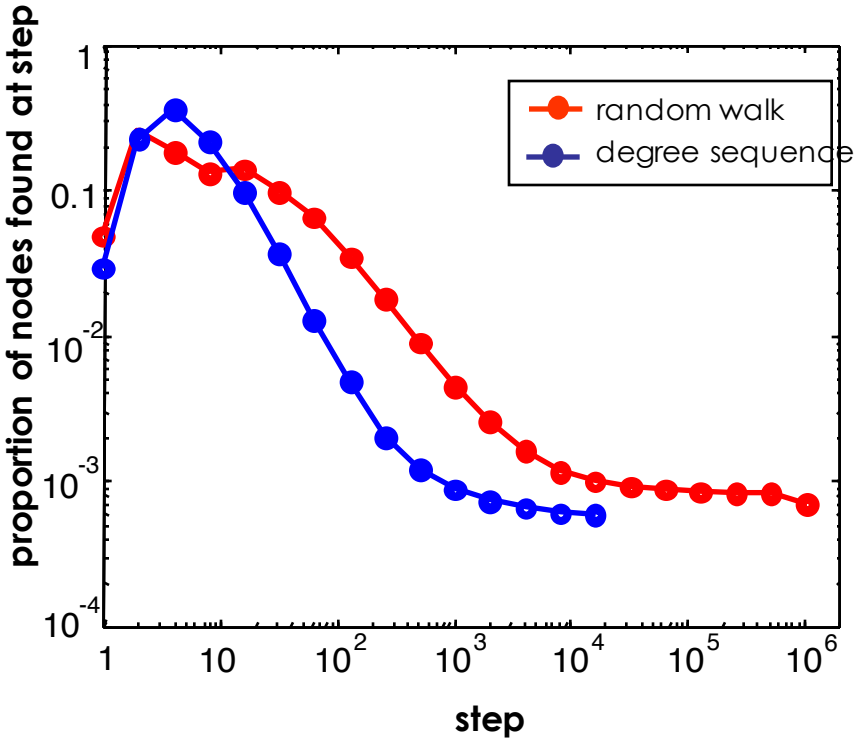
- Should not visit same node more than once

- Many neighbors of current node being visited were also neighbors of previously visited nodes, and there is a bias toward high degree nodes being 'seen' over and over again

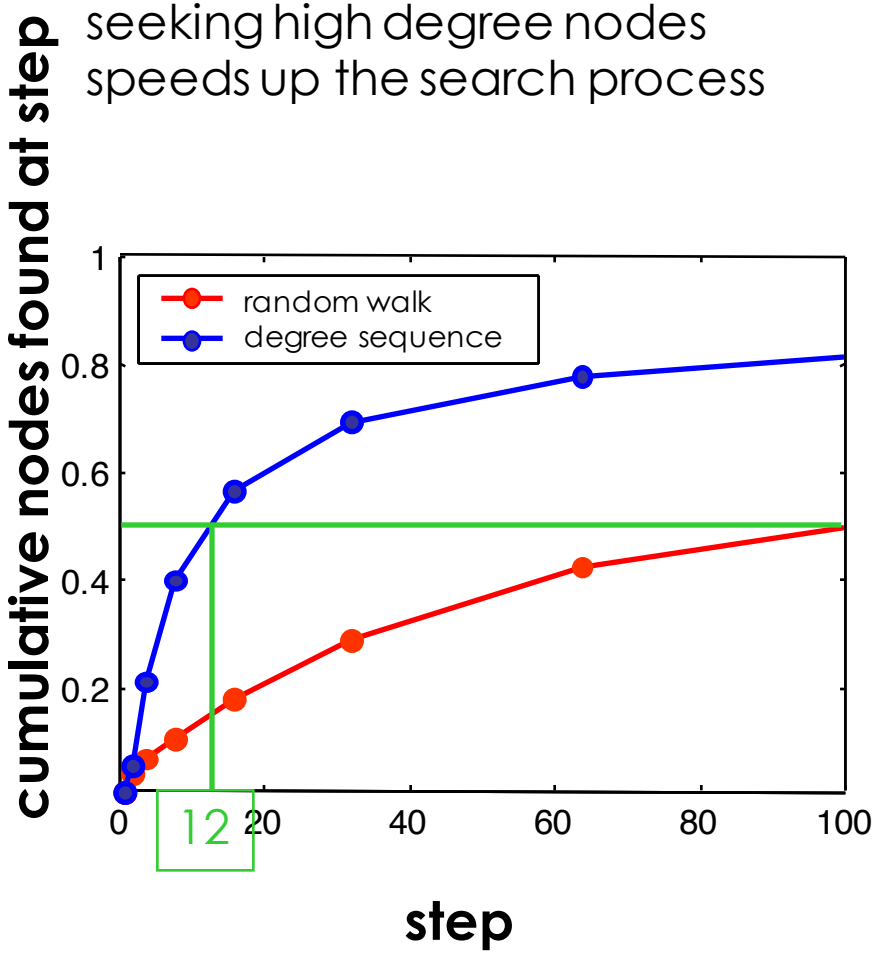
Status and degree of node visited



Progress of exploration in a 10,000 node graph knowing 2nd degree neighbors

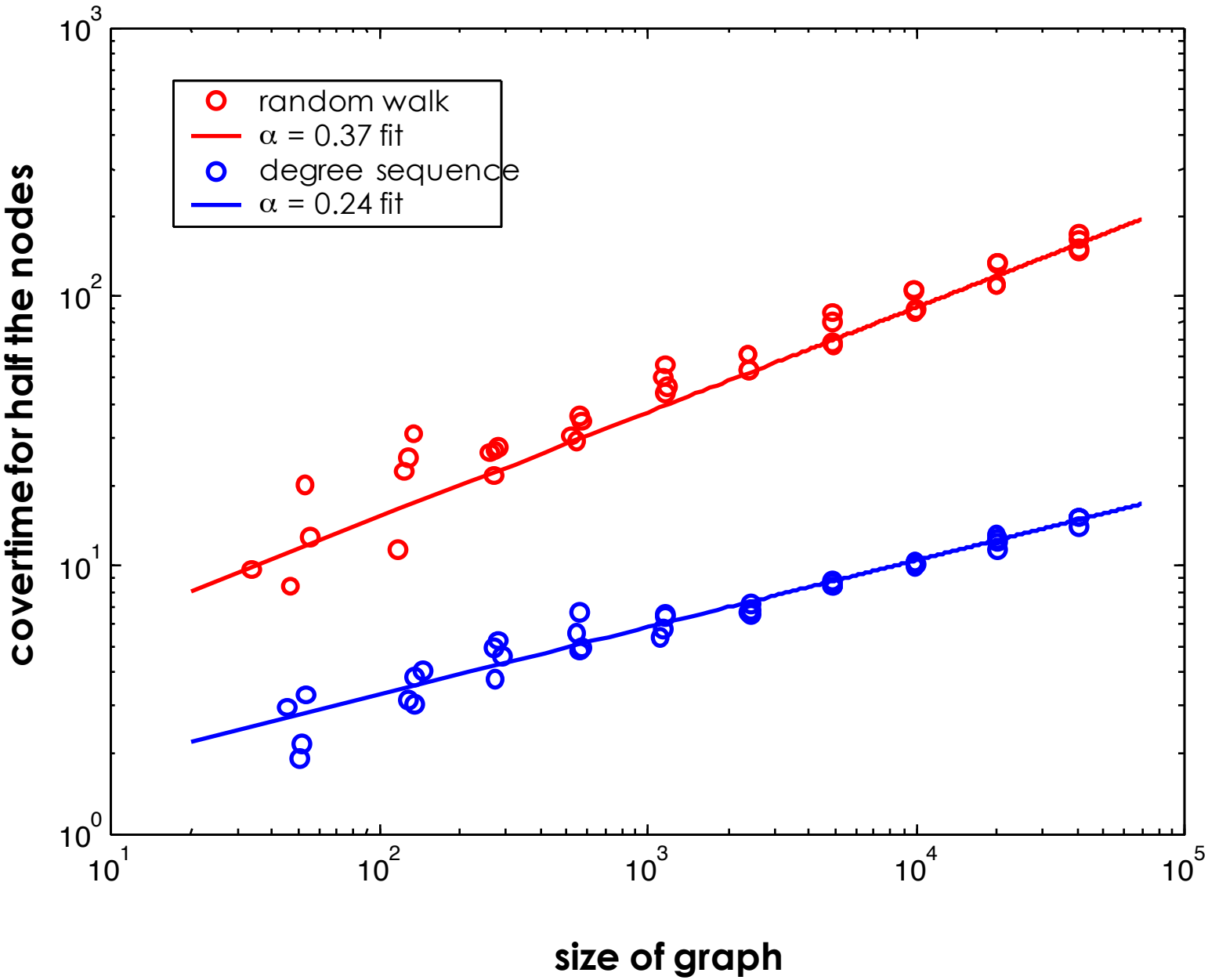


seeking high degree nodes speeds up the search process

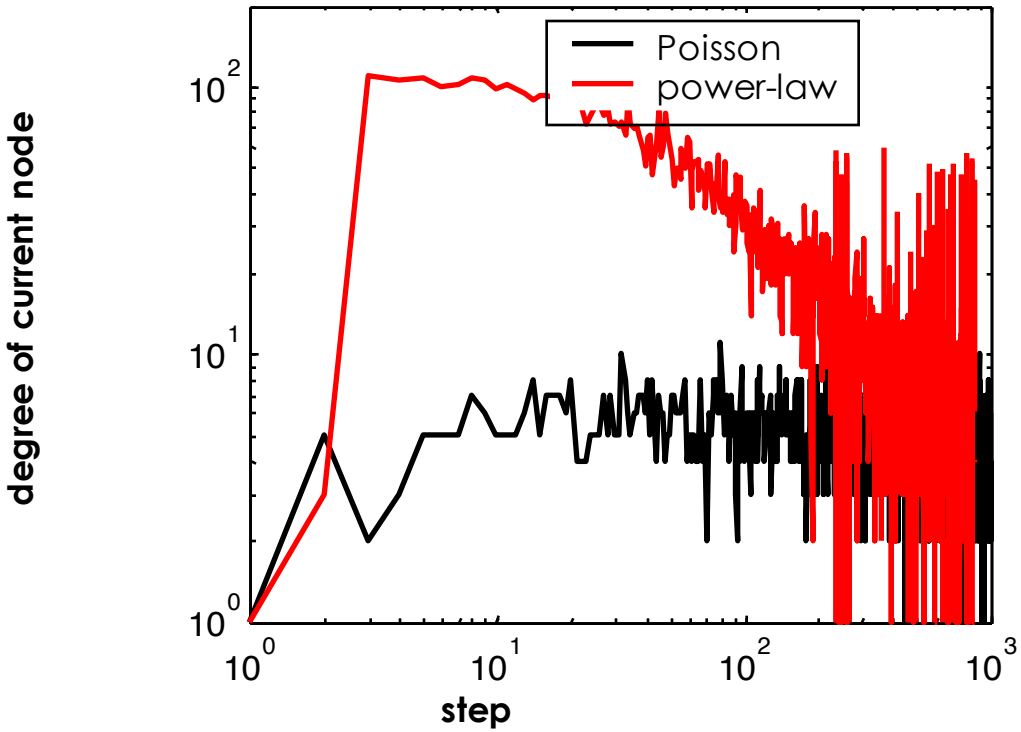


about 50% of a 10,000 node graph is explored in the first 12 steps

Scaling of search time with size of graph



Comparison with a Poisson graph

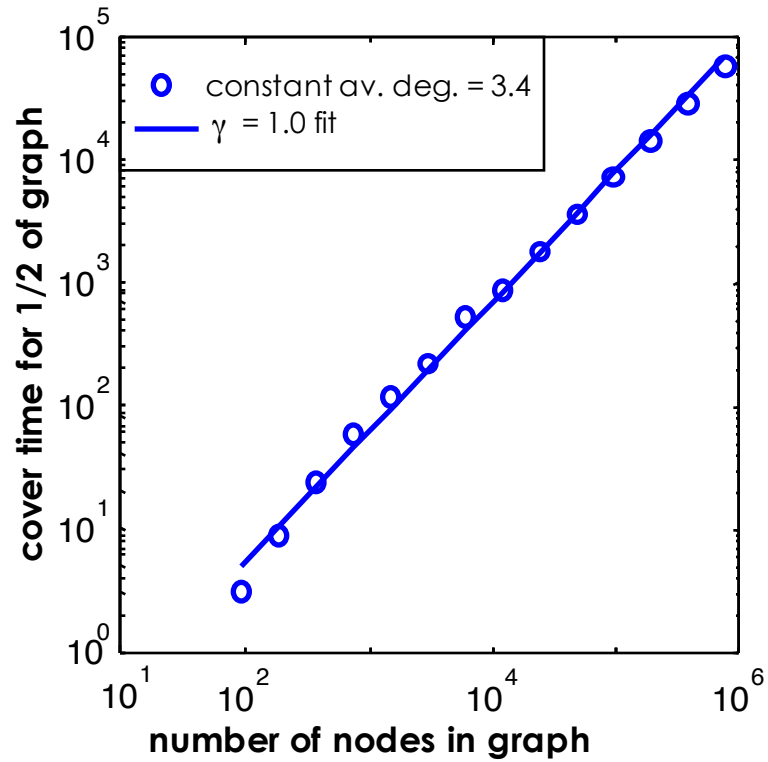


$$G_0(x) = e^{z(x-1)}$$

$$G_1(x) = \frac{x}{z} G_0'(x) = G_0(x)$$

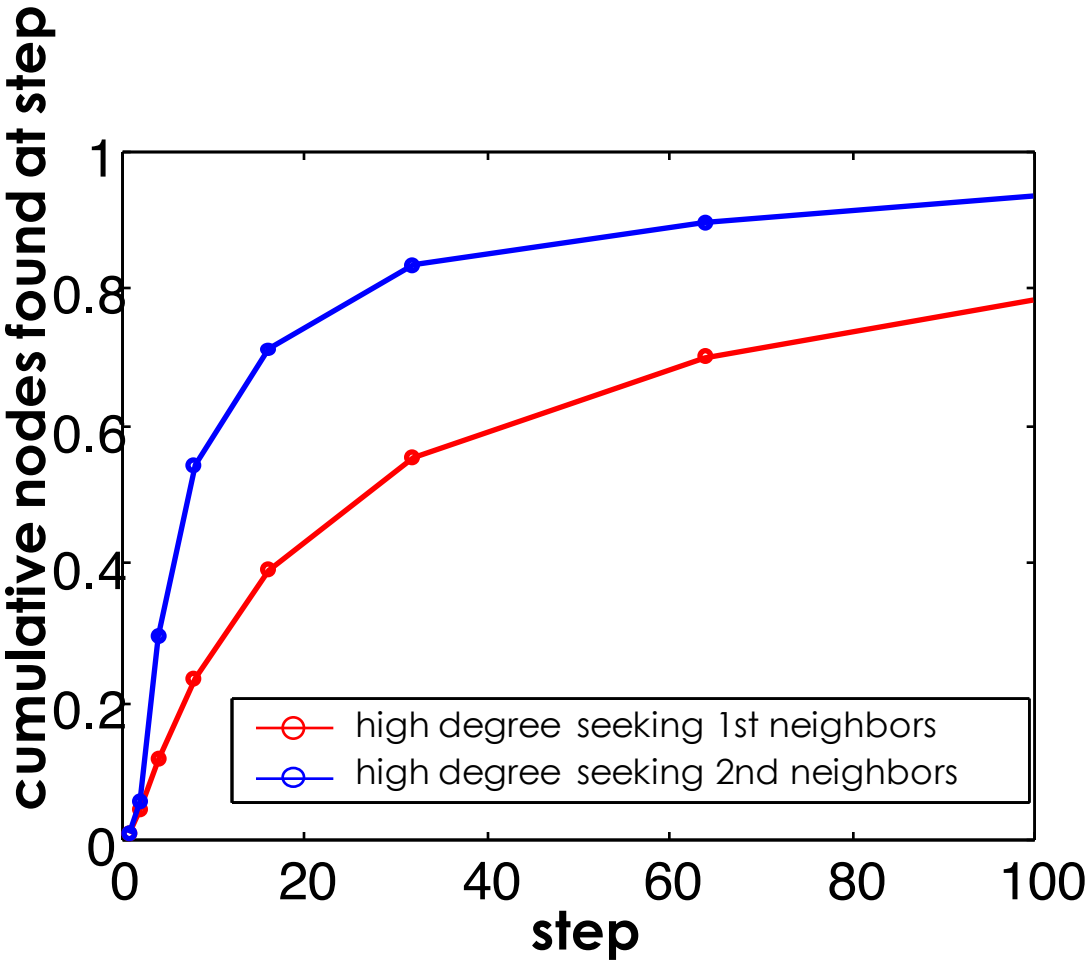
expected degree and expected degree following a link are equal

scaling is linear



Gnutella network

50% of the files in a 700 node network can be found in < 8 steps

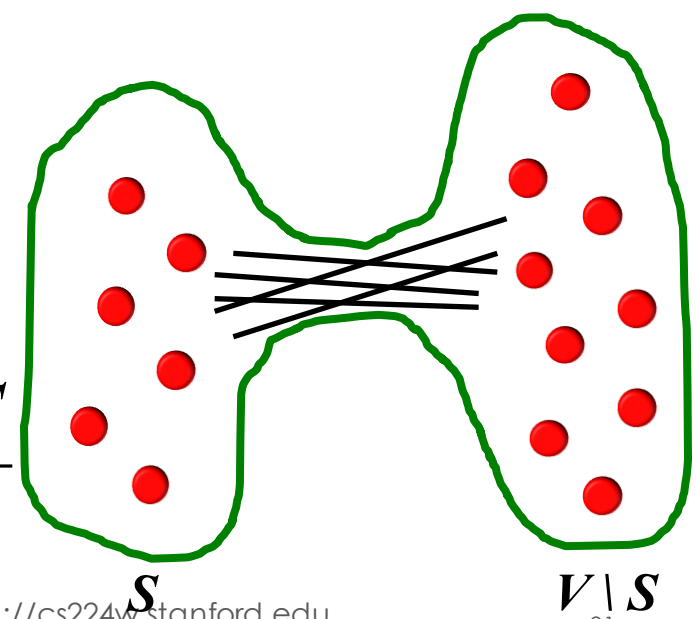


Expander graphs

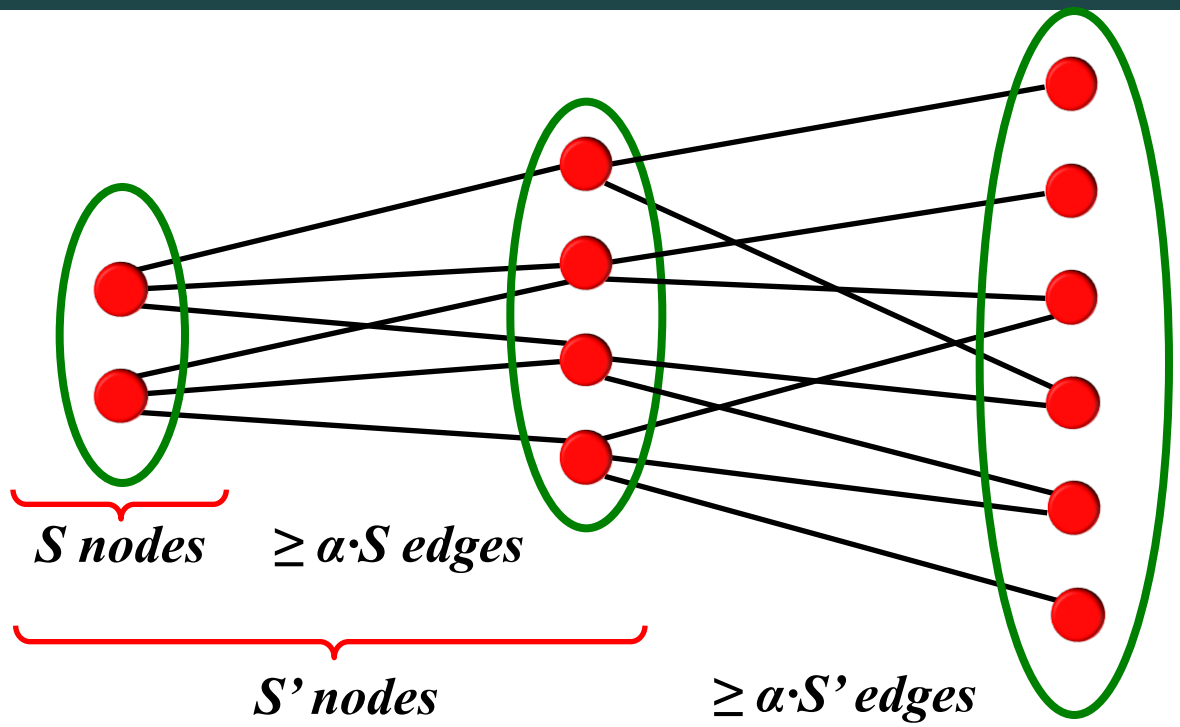
Time permitting

Def: Random k-Regular Graphs

- **We need to define two concepts**
- **1) Define: Random k-Regular graph**
  - Assume each node has k spokes (half-edges)
  - Randomly pair them up!
- **2) Define: Expansion**
  - Graph $G(V, E)$ has **expansion α**: if $\forall S \subseteq V$: #edges leaving S $\geq \alpha \cdot \min(|S|, |V \setminus S|)$
  - **Or equivalently:** $$\alpha = \min_{S \subseteq V} \frac{\#\,edges\ leaving\ S}{\min(|S|, |V \setminus S|)}$$



Expansion: Intuition



$$\alpha = \min_{S \subseteq V} \frac{\#\, edges\ leaving\ S}{\min(|S|, |V \setminus S|)}$$

(A big) graph with "good" expansion

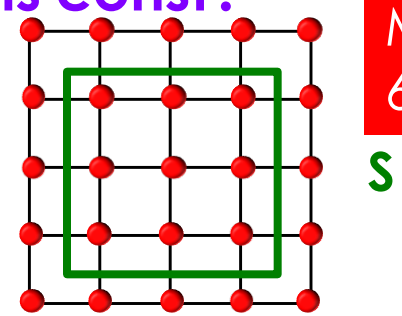
Expansion: k-Regular Graphs

$$\alpha = \min_{S \subseteq V} \frac{\#\, edges\ leaving\ S}{\min(|S|, |V \setminus S|)}$$

- ***k*-regular graph** (every node has degree k):
  - Expansion is at most k (when S is a single node)

- **Is there a graph on n nodes ($n \to \infty$), of fixed max deg. k, so that expansion α remains const?**

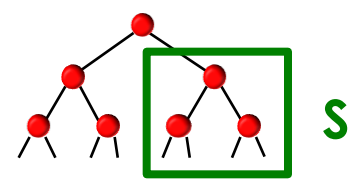
Make this into 6x6 grid!



Examples:

- **n×n grid:** $k=4$: $\alpha = 2n/(n^2/4) \to 0$
  (S=n/2 × n/2 square in the center)

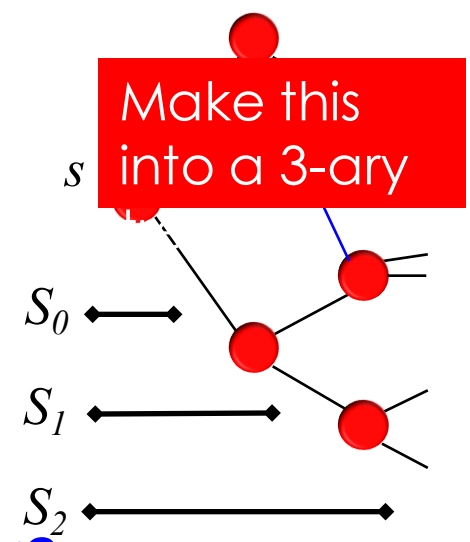
- **Complete binary tree:**
  $\alpha \to 0$ for $|S| = (n/2)-1$



- **Fact:** For a random **3-regular graph** on n nodes, there is some const α ($\alpha > 0$, independent. of n) such that w.h.p. the expansion of the graph is $\geq \alpha$ (In fact, $\alpha = d/2$ as $d \to \infty$)

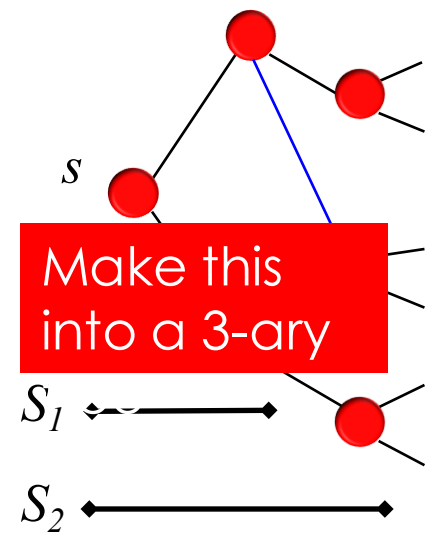
Diameter of 3-Regular Rnd. Graph

- **Fact:** In a graph on n nodes with expansion α, for all pairs of nodes s and t there is a path of $O((\log n) / \alpha)$ edges connecting them.
- Proof:
  - Proof strategy:
    - We want to show that from any node s there is a path of length $O((\log n)/\alpha)$ to any other node t
  - Let S_j be a set of all nodes found within j steps of BFS from s.
  - **How does S_j increase as a function of j?**



Diameter of 3-Regular Rnd. Graph

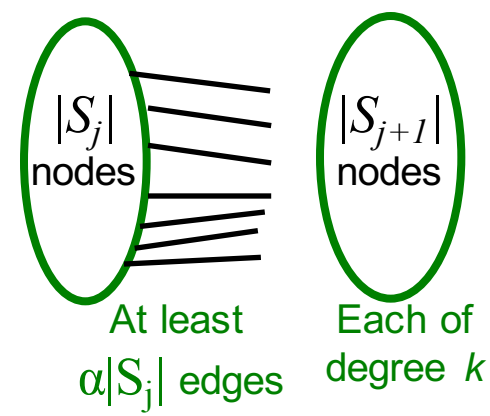
- Proof (continued):
  - Let S_j be a set of all nodes found within j steps of BFS from s.
  - **We want to relate S_j and S_{j+1}**



$$|S_{j+1}| \geq |S_j| + \frac{\overbrace{\alpha|S_j|}^{\text{Expansion}}}{\underbrace{k}_{\substack{\text{At most } k \text{ edges} \\ \text{“collide” at a node}}}} =$$

$$|S_{j+1}| \geq |S_j|\left(1+\frac{\alpha}{k}\right) = S_0\left(1+\frac{\alpha}{k}\right)^{j+1}$$

where $S_0=1$



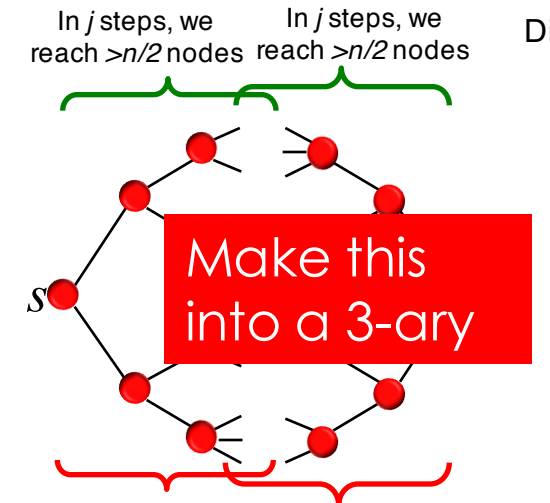
Diameter of 3-Regular Rnd. Graph

$$e = \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$$

- Proof (continued):
  - **In how many steps of BFS do we reach >*n/2* nodes?**
  - Need j so that: $S_j = \left(1+\frac{\alpha}{k}\right)^j \geq \frac{n}{2}$
  - Let's set: $j = \frac{k\log_2 n}{\alpha}$
  - Then:

$$\left(1+\frac{\alpha}{k}\right)^{\frac{k\log_2 n}{\alpha}} \geq 2^{\log_2 n} = n > \frac{n}{2}$$

  - In $2k/\alpha\cdot\log n$ steps $|S_j|$ grows to $\Theta(n)$. So, **the diameter of G is $O(\log(n)/\alpha)$**



Diameter = $2\cdot j$

Diameter = *2 log(n)*

Claim: $\left(1+\frac{\alpha}{k}\right)^{\frac{k\log_2 n}{\alpha}} \geq 2^{\log_2 n}$

Remember $n>0$, $\alpha \leq k$ then:

if $\alpha = k$: $(1+1)^{\frac{1}{1}\log_2 n} = 2^{\log_2 n}$

if $\alpha \to 0$ then $\frac{k}{\alpha} = x \to \infty$:

and $\left(1+\frac{1}{x}\right)^{x\log_2 n} = e^{\log_2 n} > 2^{\log_2 n}$

Summary

- Small world phenomenon:
  - Local structure (e.g. clustering)
  - Short average shortest path
- The Watts-Strogatz captures both
- Other models create *navigable* small-world models
- Power-law networks are navigable due to presence of hubs