

Diameter of G_{np} and the Small-World Phenomenon

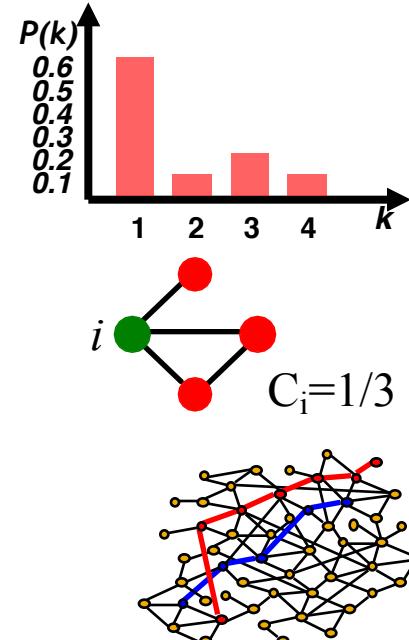
Recap: Network Properties & G_{np}

How to characterize networks?

- Connectivity
- Degree distribution $P(k)$
- Clustering Coefficient C
- Diameter (shortest path length) h

How to model networks?

- Erdös-Renyi Random Graph [Erdös-Renyi, '60]
 - $G_{n,p}$: undirected graph on n nodes where each edge (u,v) appears independently with prob. p



Random Graph Model: Edges

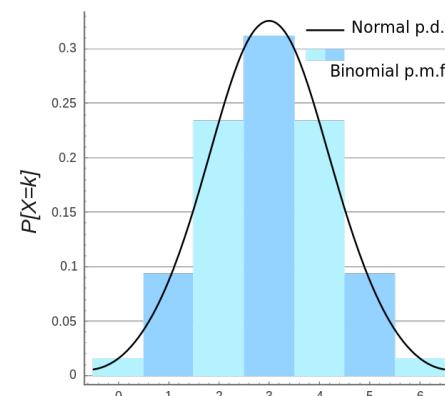
- How likely is a graph on E edges?
- $P(E)$: the probability that a given G_{np} generates a graph on exactly E edges:

$$P(E) = \binom{E_{\max}}{E} p^E (1-p)^{E_{\max}-E}$$

where $E_{\max}=n(n-1)/2$ is the maximum possible number of edges in an undirected graph of n nodes

$P(E)$ is exactly the
Binomial distribution >>>

Number of successes in a sequence of E_{\max} independent yes/no experiments



Node Degrees in a Random Graph

What is expected degree of a node?

- Let X_v be a rnd. var. measuring the degree of node v
- We want to know $E[X_v]$:

Approach:

- Decompose X_v to $X_v = X_{v,1} + X_{v,2} + \dots + X_{v,n-1}$
 - where $X_{v,u}$ is a $\{0,1\}$ -random var. which tells if edge (v,u) exists or not
- For the calculation we will need: Linearity of expectation
 - For any random variables Y_1, Y_2, \dots, Y_k
 - If $Y = Y_1 + Y_2 + \dots + Y_k$, then $E[Y] = \sum_i E[Y_i]$

$$E[X_v] = \sum_{u=1}^{n-1} E[X_{vu}] = (n-1)p$$

How to think about this?

- Prob. of node u linking to node v is p
- u can link (flips a coin) to all other $(n-1)$ nodes
- Thus, the expected degree of node u is: $p(n-1)$

Properties of G_{np}

Degree distribution: $P(k)$

Path length: h

Clustering coefficient: C

What are the values of
these properties for G_{np} ?

Degree Distribution

- **Fact:** Degree distribution of G_{np} is binomial.
- Let $P(k)$ denote the fraction of nodes with degree k :

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

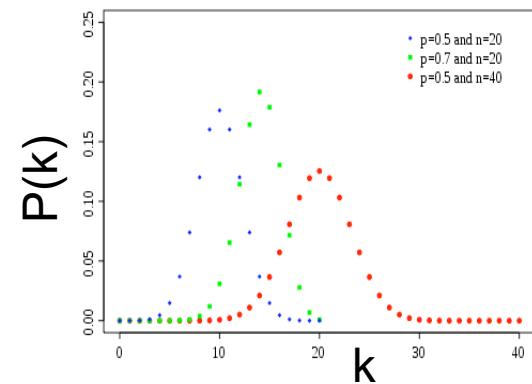
Diagram illustrating the components of the binomial probability formula:

- Select k nodes out of $n-1$ (points to the binomial coefficient term)
- Probability of having k edges (points to p^k)
- Probability of missing the rest of the $n-1-k$ edges (points to $(1-p)^{n-1-k}$)

Mean, variance of a binomial distribution

$$\bar{k} = p(n-1)$$

$$\sigma^2 = p(1-p)(n-1)$$



$$\frac{\sigma}{\bar{k}} = \left[\frac{1-p}{p} \frac{1}{(n-1)} \right]^{1/2} \approx \frac{1}{(n-1)^{1/2}}$$

By the law of large numbers, as the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of \bar{k} .

Clustering Coefficient of G_{np}

- **Remember:** $C_i = \frac{2e_i}{k_i(k_i - 1)}$ Where e_i is the number of edges between i's neighbors
- Edges in G_{np} appear i.i.d. with prob. p
- **So:** $e_i = p \frac{k_i(k_i - 1)}{2}$
 - Each pair is connected with prob. p
 - Number of distinct pairs of neighbors of node i of degree k_i
- **Then (expected):** $C = \frac{p \cdot k_i(k_i - 1)}{k_i(k_i - 1)} = p = \frac{\bar{k}}{n-1} \approx \frac{\bar{k}}{n}$

Clustering coefficient of a random graph is small.

If we generate bigger and bigger graphs with fixed avg. degree k (that is we set $p = k \cdot 1/n$), then C decreases with the graph size n .

Network Properties of G_{np}

Degree distribution:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Clustering coefficient:

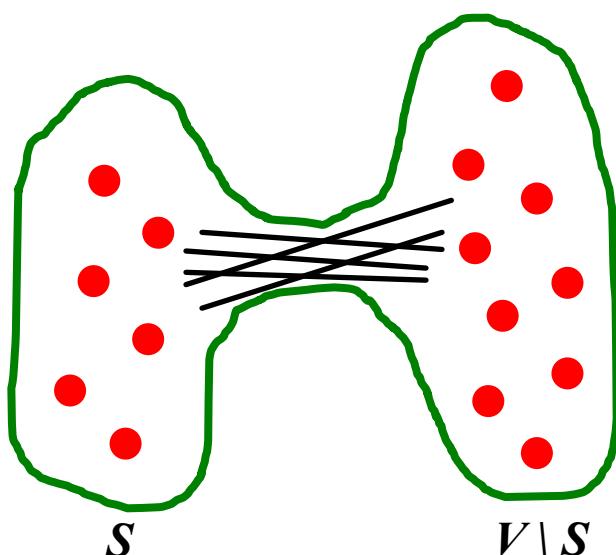
$$C = p = \bar{k}/n$$

Path length:

next!

Two Definitions

- 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** :
$$\text{if } \forall S \subseteq V: \# \text{edges leaving } S \geq \alpha \cdot \min(|S|, |V \setminus S|)$$

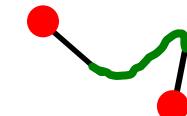


Def 1 : Random k-Regular Graphs

- To prove the diameter of a G_{np} we define a few concepts
- **Define: Random k-Regular graph**

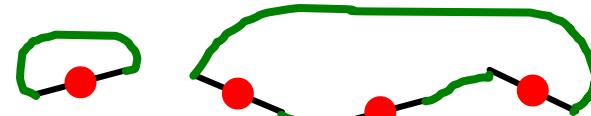
- Assume each node has k spokes (half-edges)

■ $k=1$:



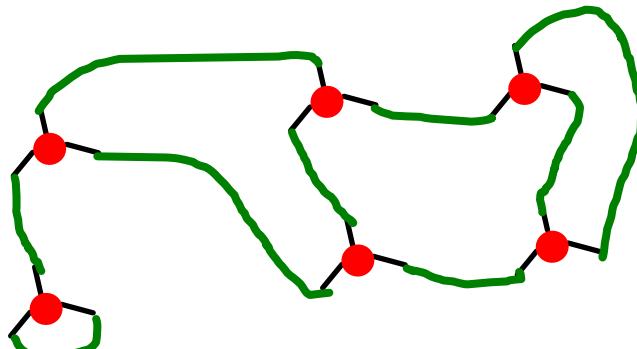
Graph is a set of pairs

■ $k=2$:



Graph is a set of cycles

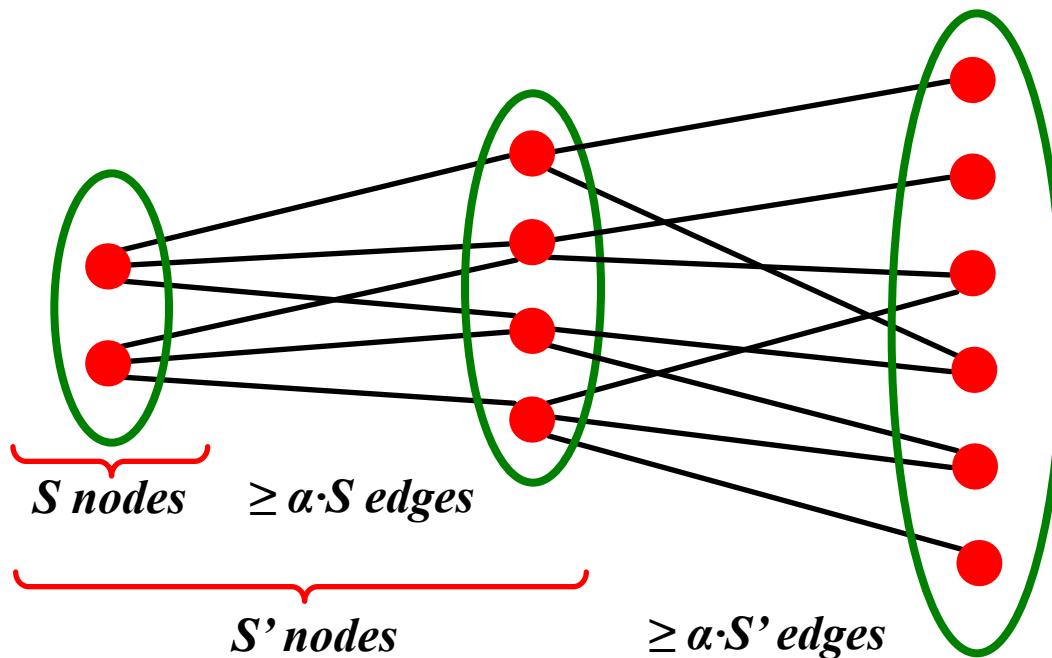
■ $k=3$:



Arbitrarily complicated graphs

- Randomly pair them up!

Def 2: Expansion: Intuition

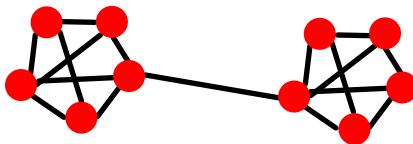
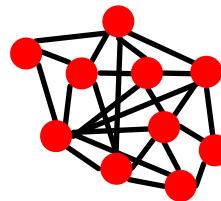
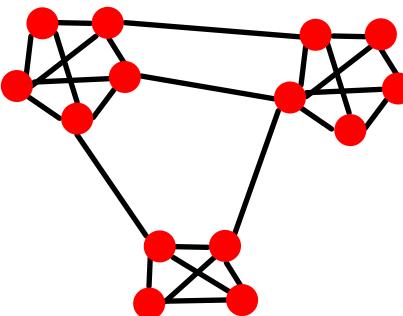


$$\alpha = \min_{S \subseteq V} \frac{\#\text{edges leaving } S}{\min(|S|, |V \setminus S|)}$$

(A big) graph with “good” expansion

Expansion: Measures Robustness

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

- Expansion is **measure of robustness**:
 - To disconnect l nodes, we need to cut $\geq \alpha \cdot l$ edges
- Low expansion:
- High expansion:
- Social networks:
 - “Communities”

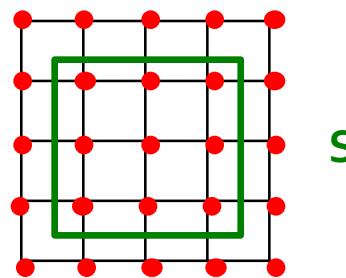
Def 2: Expansion of k-Regular Graphs

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

- **k -regular graph:**
 - Expansion is at most k (when S is a single node)
- Is there a graph on n nodes ($n \rightarrow \infty$), of fixed max deg. k , so that expansion α remains const?

Examples:

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



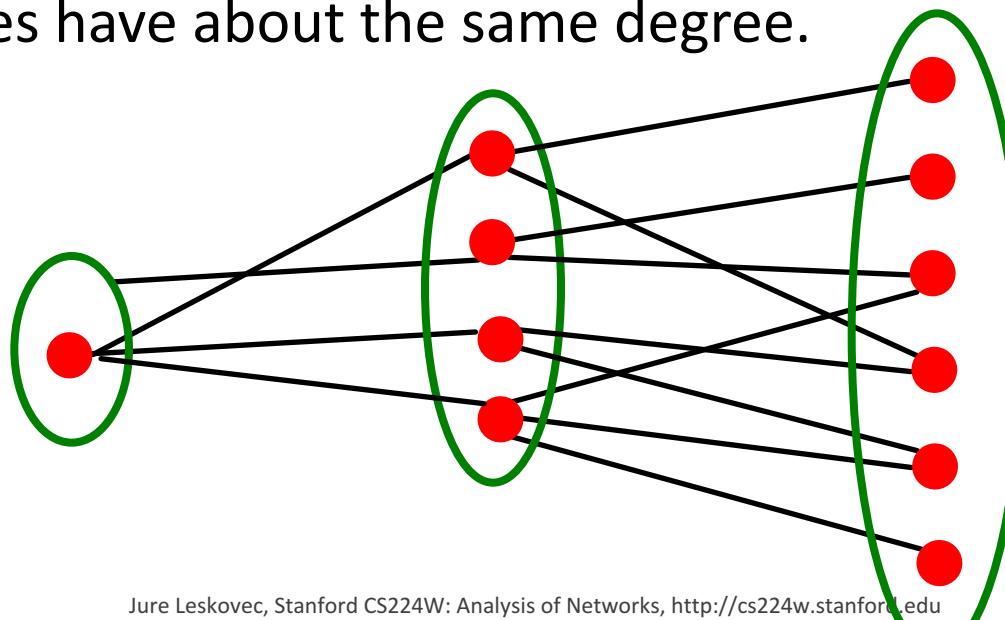
Let $m \leq n/2$ nodes on one side. Consider the $3m$ spokes out of them. Then, $\alpha = 3m[(n-m)/n]/m \geq 3/2$.

- **Fact:** For a random **3-regular graph** on n nodes, there is some const α ($\alpha > 0$, independent. of n) such that with high prob. (prob $\rightarrow 1$ as $n \rightarrow \infty$) the expansion of the graph is $\geq \alpha$.

In fact, $\alpha = \Theta(k/2)$ as $n \rightarrow \infty$

Diameter of k-Regular Rnd. Graph

- We will prove the diameter of a k -regular random graph ($k=3$)
 - Note that k -regular random graph are essentially the same as random graphs
 - In a random graph variance/mean degree goes to 0 as graph size increases, which means that intuitively all nodes have about the same degree.

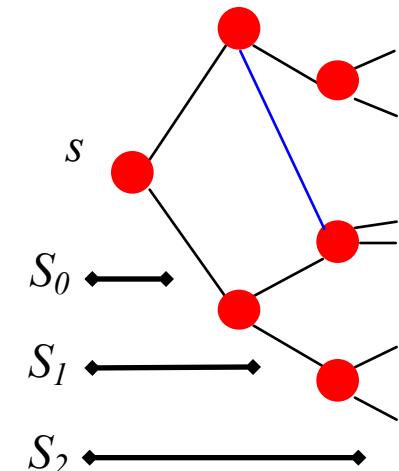


Diameter of k -Regular Rnd. Graph

Fact: In a k -regular random 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 k-Regular Rnd. Graph

Proof (continued):

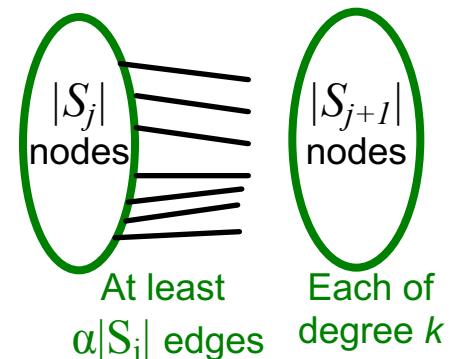
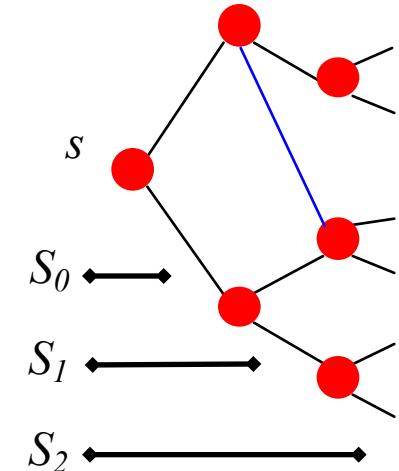
- Let S_j be a set of all nodes found within j steps of BFS from s .
- We want to relate S_j and S_{j+1}

$$|S_{j+1}| \geq |S_j| + \frac{\alpha |S_j|}{k} =$$

Expansion
 $\overbrace{}^{\alpha |S_j|}$
 $\underbrace{}_{\text{At most } k \text{ edges "collide" at a node}}$

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

where $S_0 = 1$



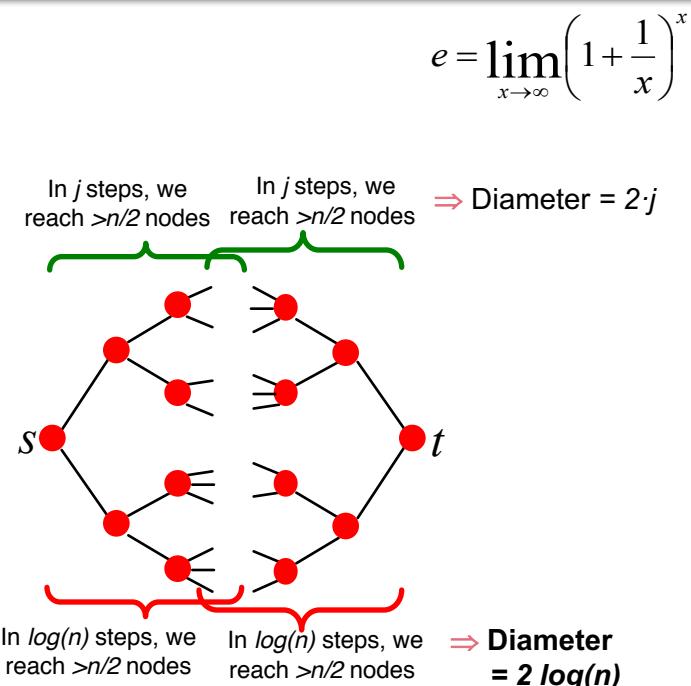
Diameter of k-Regular Rnd. Graph

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 $O(\log(n)/\alpha)$



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:

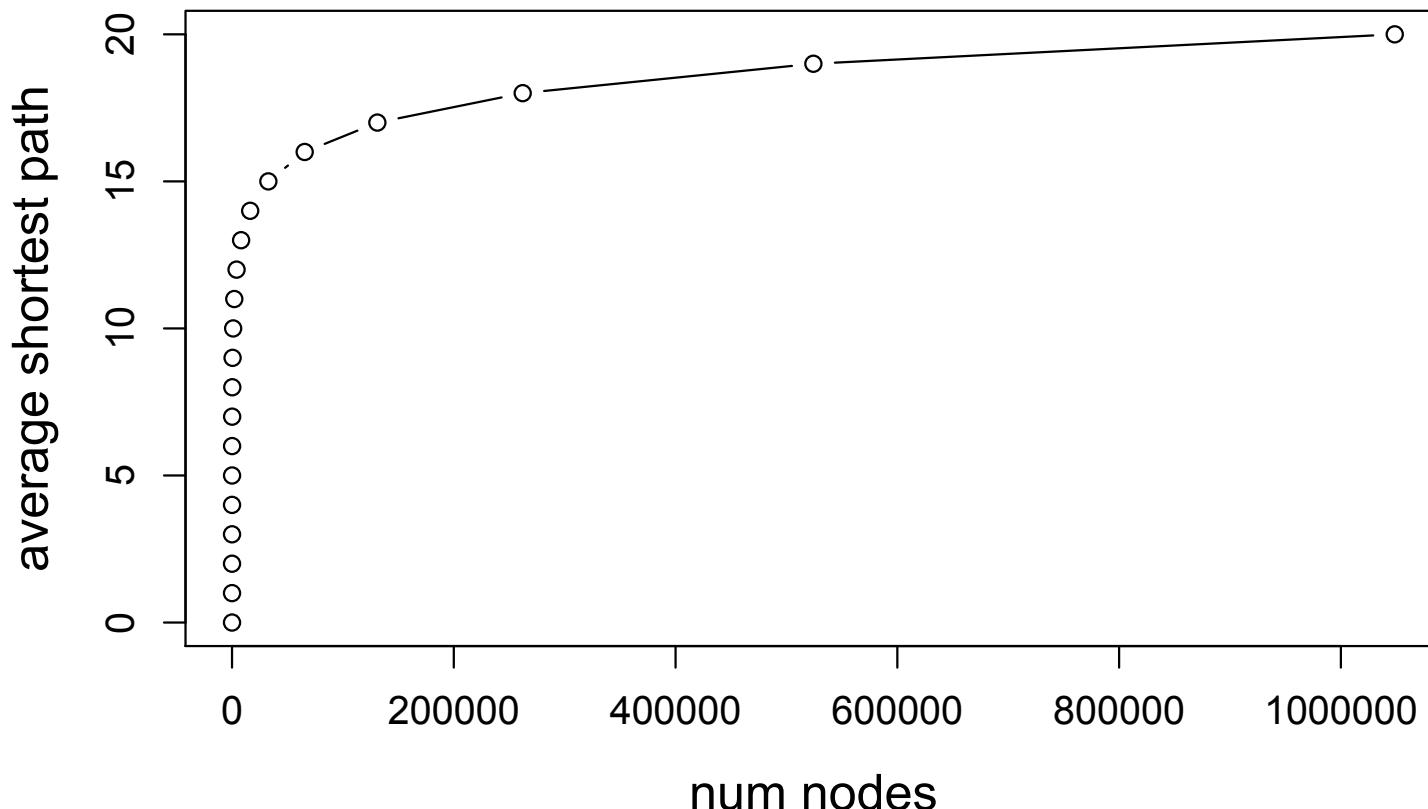
if $\alpha = k : (1+1)^{\frac{1}{1} \log_2 n} = 2^{\log_2 n}$

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

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

Erdös-Renyi avg. shortest path

- Erdös-Renyi networks can grow to be very large but nodes will be just a few hops apart



Back to Node Degrees of G_{np}

Extra

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

- Remember, expected degree: $E[X_v] = (n-1)p$
- To have const. degree we want $E[X_v]$ be independent of n : So let: $p=k/(n-1)$
- Observation: If we build random graph G_{np} with $p=k/(n-1)$ we have many isolated nodes
- Why?

$$P[v \text{ has degree } 0] = (1-p)^{n-1} = \left(1 - \frac{k}{n-1}\right)^{n-1} \xrightarrow[n \rightarrow \infty]{} e^{-k}$$

Why?

$$\lim_{n \rightarrow \infty} \left(1 - \frac{k}{n-1}\right)^{n-1} = \left(1 - \frac{1}{x}\right)^{x \cdot k} = \left[\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^{-x} \right]^{-k} = e^{-k}$$

Use substitution $\frac{1}{x} = \frac{k}{n-1}$

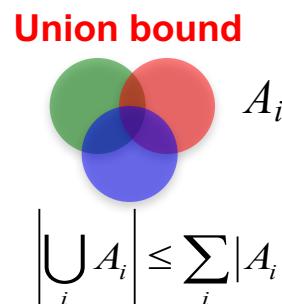
In other words: In the limit of $n \rightarrow \infty$ Poisson distribution is a good approximation of a Binomial distribution and we computed $P(x=0)$ for the Poisson PMF.

Poisson dist: $\Pr(X=k) = (\lambda^k * \exp(-\lambda)) / k!$. It is #event occurrence in a time interval, where rate is constant λ . Here, $k=0$, $\lambda=k$.

No Isolated Nodes

- How big do we have to make p before we are likely to have no isolated nodes?
- We know: $P[v \text{ has degree } 0] = e^{-k}$
- Event we are asking about is:
 - $I = \text{some node is isolated}$
 - $I = \bigcup_{v \in N} I_v$ where I_v is the event that v is isolated

$$P(I) = P\left(\bigcup_{v \in N} I_v\right) \leq \sum_{v \in N} P(I_v) = n e^{-k}$$



No Isolated Nodes

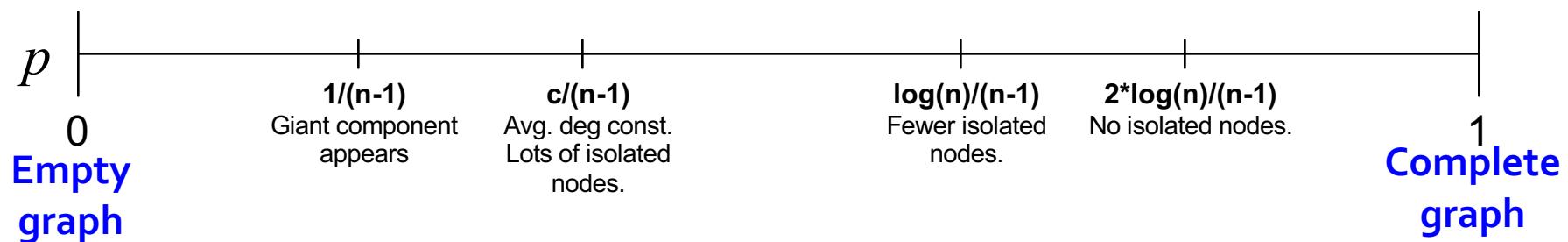
- We just learned: $P(I) \leq n e^{-k}$
- Let's try:
 - $k = \ln n$ then: $n e^{-k} = n e^{-\ln n} = n \cdot 1/n = 1$
 - $k = 2 \ln n$ then: $n e^{-2 \ln n} = n \cdot 1/n^2 = 1/n$
- So if:
 - $k = \ln n$ then: $P(I) \leq 1$
 - $k = 2 \ln n$ then: $P(I) \leq 1/n \rightarrow 0 \text{ as } n \rightarrow \infty$

So, for $p=2\ln(n)$ we get no isolated nodes
(as $n \rightarrow \infty$)

“Evolution” of a Random Graph

Extra

- Graph structure of G_{np} as p changes:

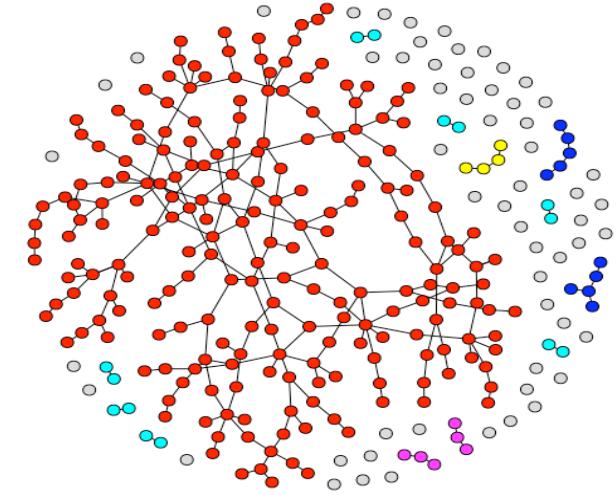
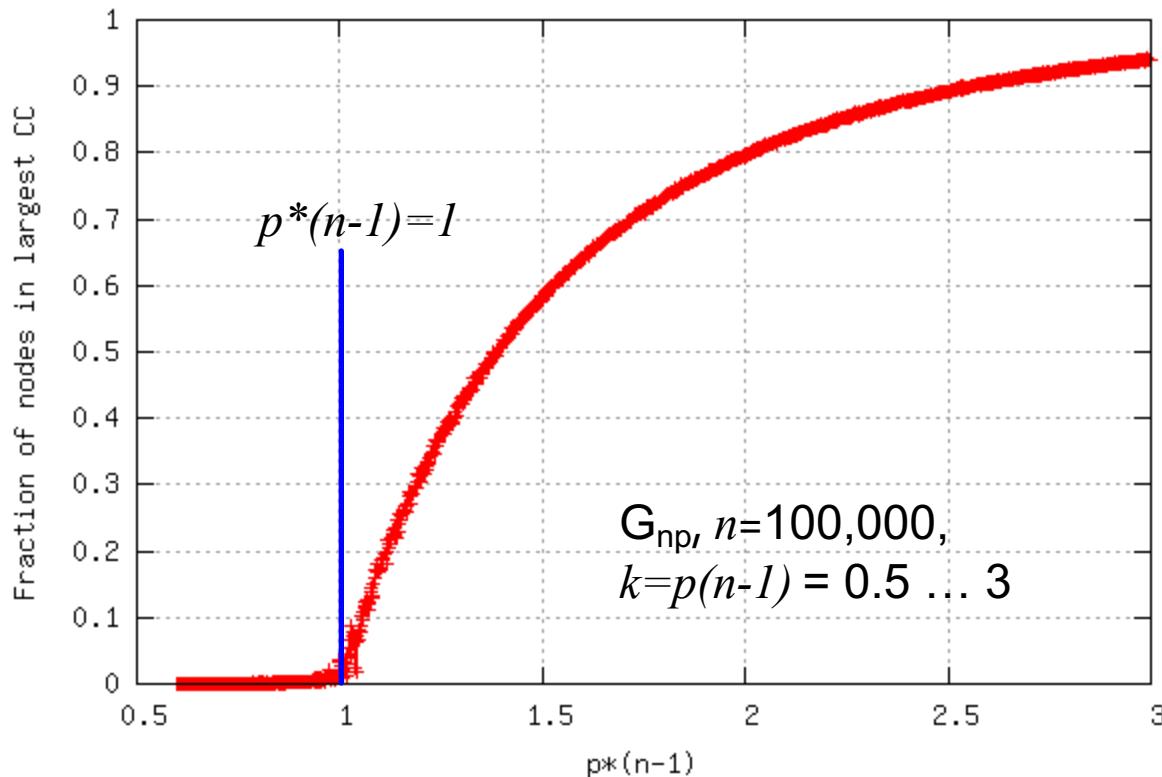


- Emergence of a giant component:

avg. degree $k=2E/n$ or $p=k/(n-1)$

- $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
- $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

G_{np} Simulation Experiment



Fraction of nodes in the largest component

- **Emergence of a giant component:**
avg. degree $k=2E/n$ or $p=k/(n-1)$
 - $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
 - $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

Network Properties of G_{np}

Degree distribution:

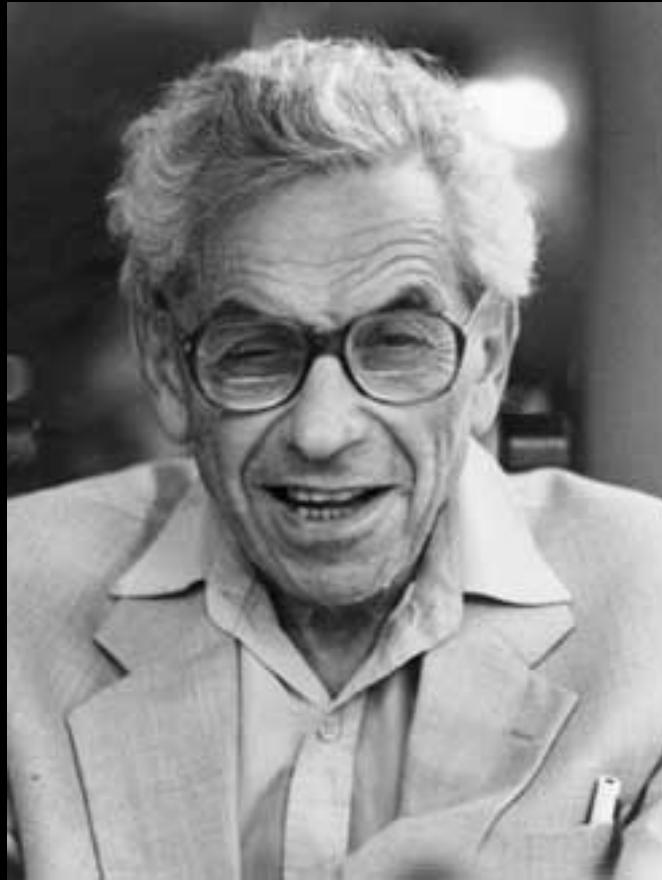
$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Path length:

$$O(\log n)$$

Clustering coefficient:

$$C = p = \bar{k} / n$$



Paul Erdős

G_{np} is so cool!
Let's compare it to real networks.

**DIRECT FROM
★ RINGSIDE! ★
THE FIGHT EVERYONE WANTS TO SEE...**

**MSN
BOMBER**

VS
**Gnp
ROCKER**

**ROCK'EM
SOCK'EM**
ROBOTS™



MAIN EVENT

Back to MSN vs. G_{np}

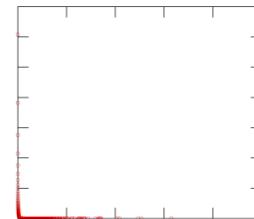
Degree distribution:

Path length:

Clustering coefficient:

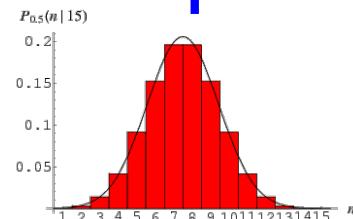
Connected component:

MSN



G_{np}

$n=180M$



6.6

$O(\log n)$



$h \approx 8.2$

\bar{k} / n



$C \approx 8 \cdot 10^{-8}$

GCC exists
when $\bar{k} > 1$.
 $\bar{k} \approx 14$.



Real Networks vs. G_{np}

- **Are real networks like random graphs?**
 - Average path length: ☺
 - Giant connected component: ☺
 - Clustering Coefficient: ☹
 - Degree Distribution: ☹
- **Problems with the random network model:**
 - Degree distribution differs from that of real networks
 - Giant component in most real networks does NOT emerge through a phase transition
 - No “local” structure – clustering coefficient is too low
- **Most important: Are real networks random?**
 - The answer is simply: **NO!**

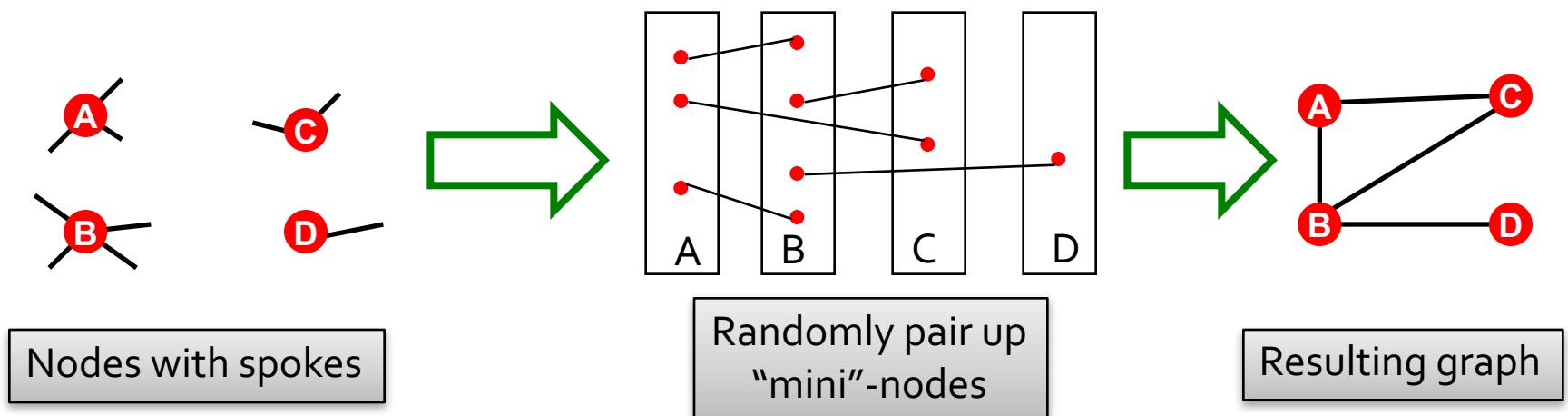
Real Networks vs. G_{np}

- If G_{np} is wrong, why did we spend time on it?
 - It is the reference model for the rest of the class
 - It will help us calculate many quantities, that can then be compared to the real data
 - It will help us understand to what degree is a particular property the result of some random process

So, while G_{np} is WRONG, it will turn out to be extremely USEFUL NULL MODEL!

Intermezzo: Configuration Model

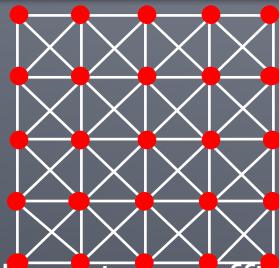
- **Goal:** Generate a random graph with a given degree sequence $k_1, k_2, \dots k_N$
- **Configuration model:**



- **Useful as a “null” model of networks:**
 - We can compare the real network G and a “random” G' which has the same degree sequence as G

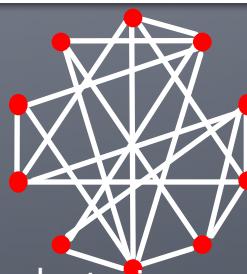
The Small-World Model

Can we have high clustering while also having short paths?



High clustering coefficient,
High diameter

Vs.

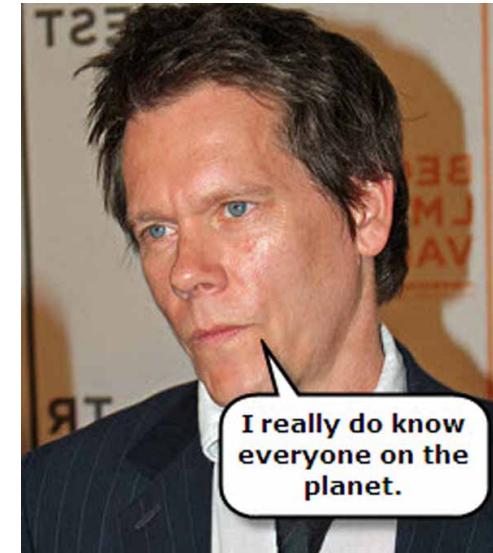


Low clustering coefficient
Low diameter

Six Degrees of Kevin Bacon

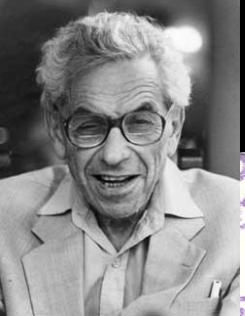
Origins of a small-world idea:

- **The Bacon number:**
 - Create a network of Hollywood actors
 - Connect two actors if they co-appeared in the movie
 - **Bacon number:** number of steps to Kevin Bacon
- As of Dec 2007, the highest (finite) Bacon number reported is 8
- Only approx. 12% of all actors cannot be linked to Bacon



Elvis Presley has a Bacon number of 2.





Erdős numbers are small!
(Jure's number is 3.)

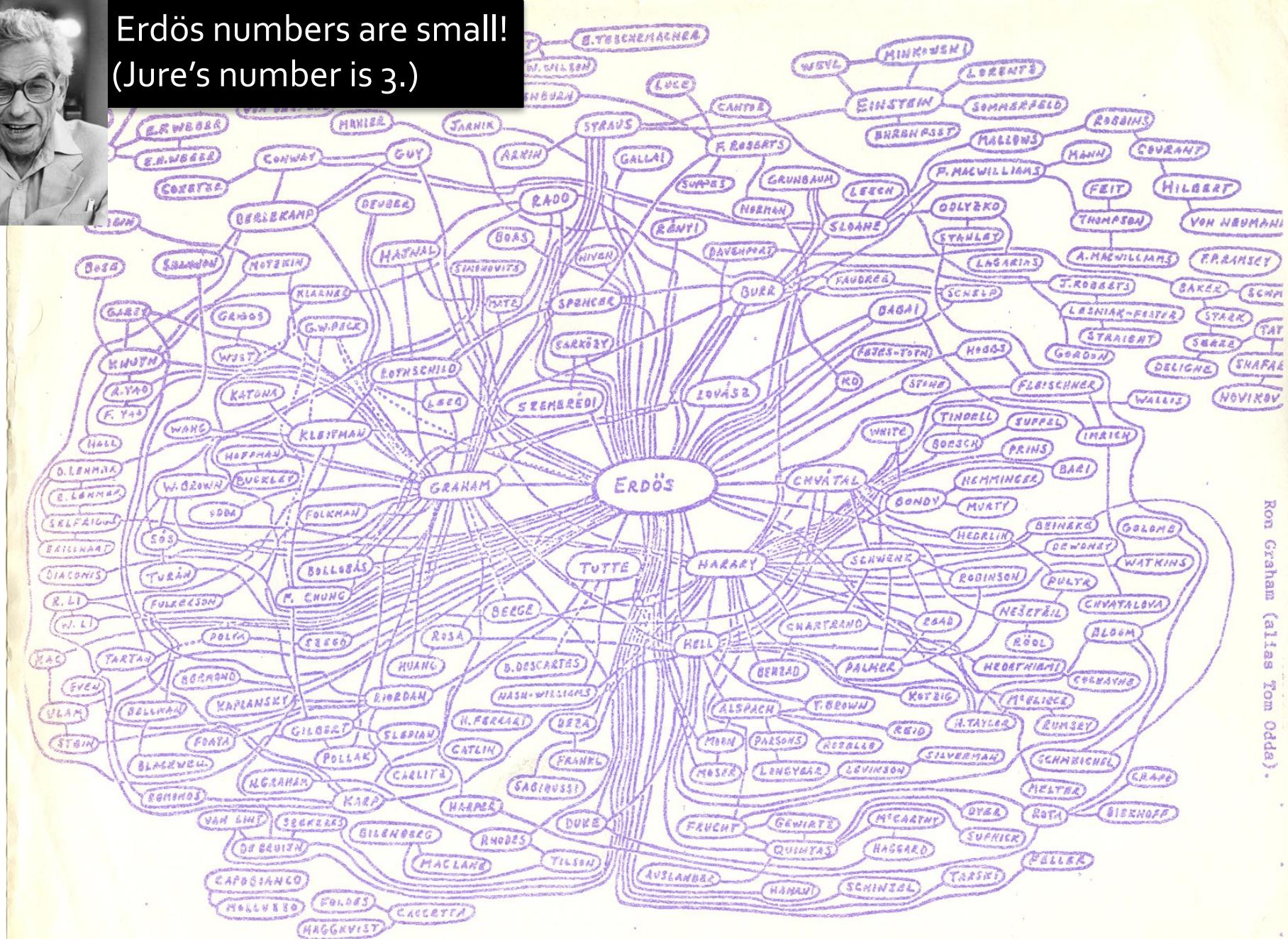
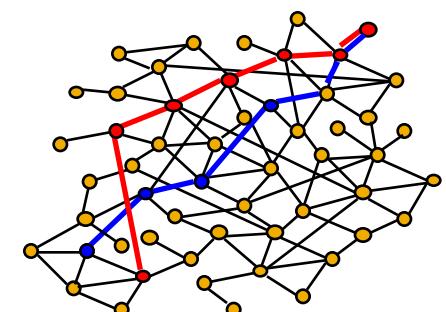


Figure 1
To appear in Topics in Graph Theory (P. Harary, ed.) New York Academy of Sciences (1979).

Find out your Erdos number: <http://www.ams.org/mathscinet/collaborationDistance.html>

The Small-World Experiment

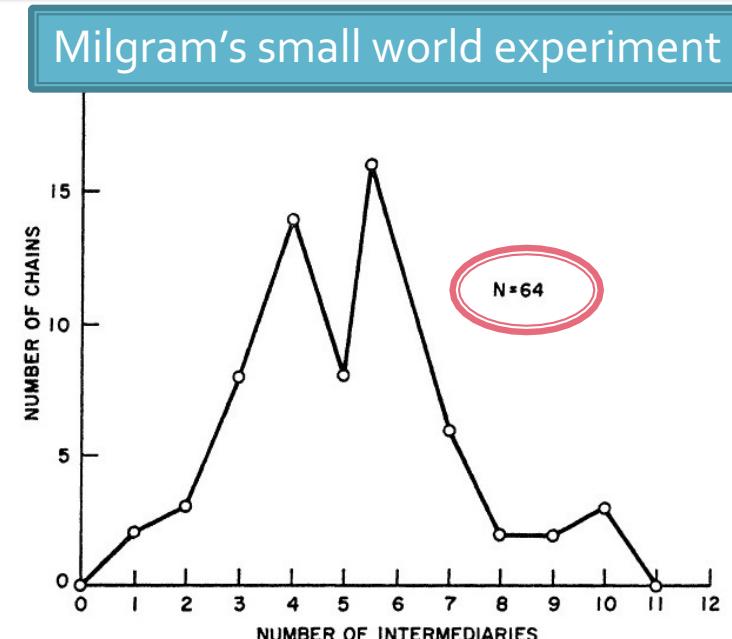
- What is the typical shortest path length between any two people?
 - Experiment on the global friendship network
 - Can't measure, need to probe explicitly
- Small-world experiment [Milgram '67]
 - Picked 300 people in Omaha, Nebraska and Wichita, Kansas
 - Ask them to get a letter to a stock-broker in Boston by passing it through friends
- How many steps did it take?



The Small-World Experiment

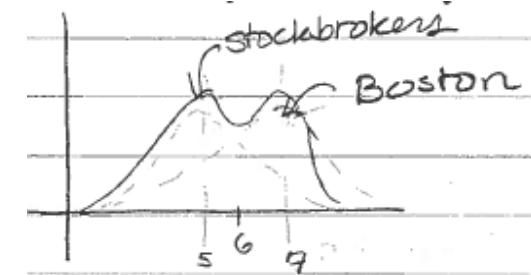
- **64 chains completed:**
(i.e., 64 letters reached the target)
 - It took 6.2 steps on the average, thus
“6 degrees of separation”

- **Further observations:**
 - People who owned stock had shorter paths to the stockbroker than random people: 5.4 vs. 6.7
 - People from the Boston area have even closer paths: 4.4



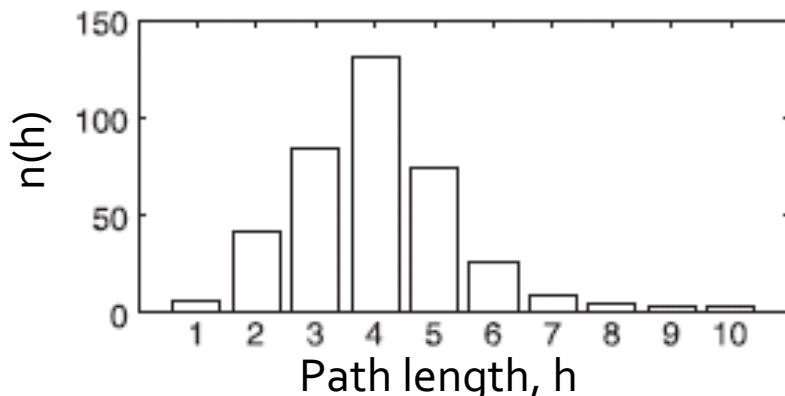
Milgram: Further Observations

- **Boston vs. occupation networks:**
- **Criticism:**
 - **Funneling:**
 - 31 of 64 chains passed through 1 of 3 people as their final step → **Not all links/nodes are equal**
 - Starting points and the target were non-random
 - There are not many samples (only 64)
 - People refused to participate (25% for Milgram)
 - Not all searches finished (only 64 out of 300)
 - **Some sort of social search:** People in the experiment follow some strategy instead of forwarding the letter to everyone. **They are not finding the shortest path!**
 - People might have used extra information resources



Columbia Small-World Study

- In 2003 Dodds, Muhamad and Watts performed the experiment using e-mail:
 - 18 targets of various backgrounds
 - 24,000 first steps ($\sim 1,500$ per target)
 - 65% dropout per step
 - 384 chains completed (1.5%)



Avg. chain length = 4.01

Problem: People stop participating

$$\text{Correction factor: } n^*(h) = \frac{n(h)}{\prod_{i=0}^{h-1} (1 - r_i)}$$

r_i drop-out rate at hop i

Small-World in Email Study

- After the correction:

- Typical path length $h = 7$

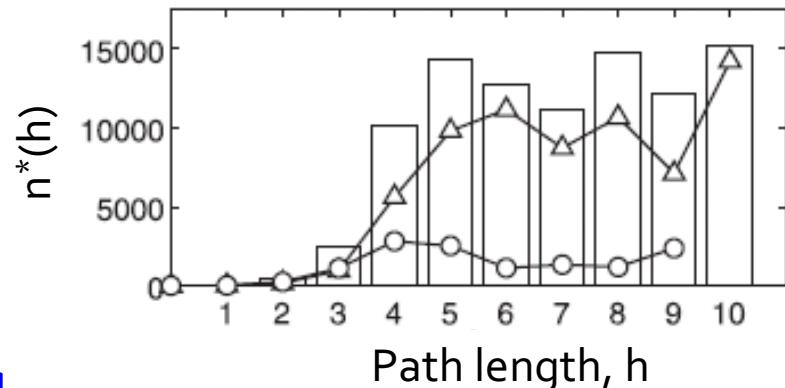
- Some not well understood phenomena in social networks:

- **Funneling effect:** Some target's friends are more likely to be the final step
 - Conjecture: High reputation/authority

- **Effects of target's characteristics:**

Structurally, why are high-status targets easier to find?

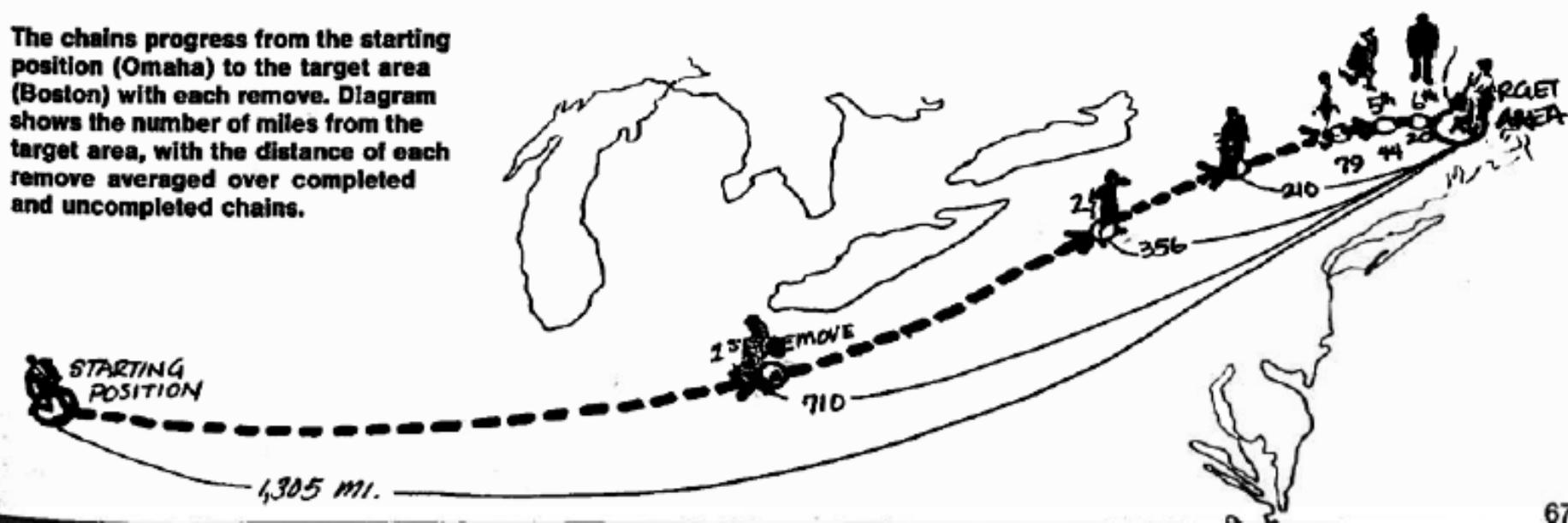
- Conjecture: Core-periphery network structure



Two Questions

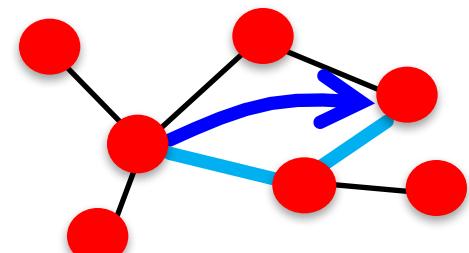
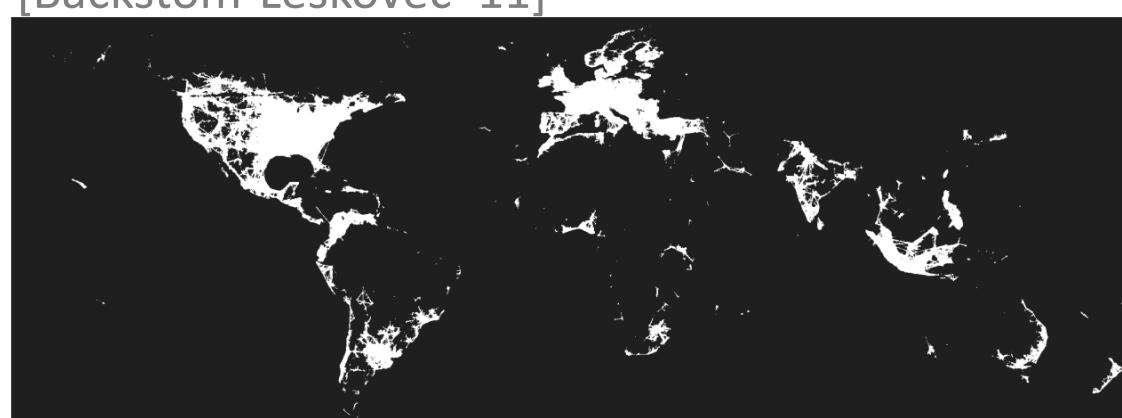
- (1) What is the structure of a social network?
- (2) What kind of mechanisms do people use to route and find the target?

The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.



6-Degrees: Should We Be Surprised?

- Assume each human is connected to 100 other people
Then:
 - Step 1: reach 100 people
 - Step 2: reach $100 * 100 = 10,000$ people
 - Step 3: reach $100 * 100 * 100 = 1,000,000$ people
 - Step 4: reach $100 * 100 * 100 * 100 = 100M$ people
 - In 5 steps we can reach 10 billion people
- What's wrong here?
 - 92% of new FB friendships are to a friend-of-a-friend
[Backstrom-Leskovec '11]



Clustering Implies Edge Locality

- MSN network has 7 orders of magnitude larger clustering than the corresponding G_{np} !
- Other examples:

Actor Collaborations (IMDB): $N = 225,226$ nodes, avg. degree $\bar{k} = 61$

Electrical power grid: $N = 4,941$ nodes, $\bar{k} = 2.67$

Network of neurons: $N = 282$ nodes, $\bar{k} = 14$

Network	h_{actual}	h_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power Grid	18.70	12.40	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

h ... Average shortest path length

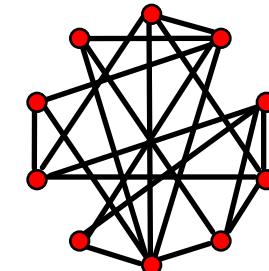
C ... Average clustering coefficient

“actual” ... real network

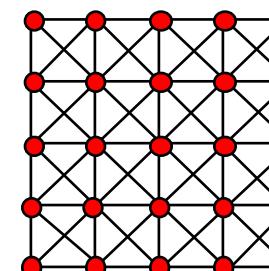
“random” ... random graph with same avg. degree

The “Controversy”

- **Consequence of expansion:**
 - **Short paths: $O(\log n)$**
 - This is the smallest diameter we can get if we have a constant degree.
 - But clustering is low!
- **But networks have “local” structure:**
 - **Triadic closure:**
Friend of a friend is my friend
 - High clustering but diameter is also high
- **How can we have both?**



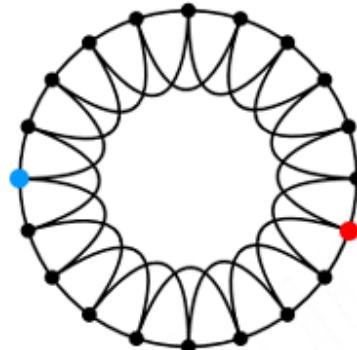
Low diameter
Low clustering coefficient



High clustering coefficient
High diameter

Small-World: How?

- Could a network with high clustering be at the same time a small world?
 - How can we at the same time have **high clustering** and **small diameter**?



High clustering
High diameter



Low clustering
Low diameter

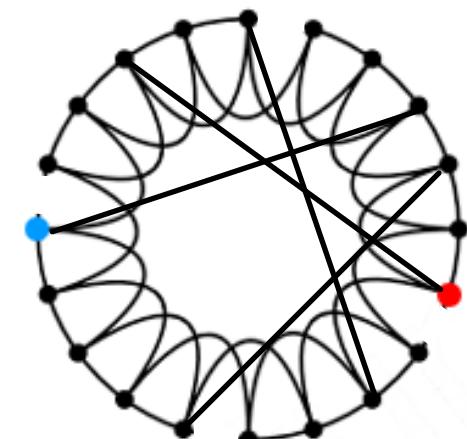
- Clustering implies edge “locality”
- Randomness enables “shortcuts”

Solution: The Small-World Model

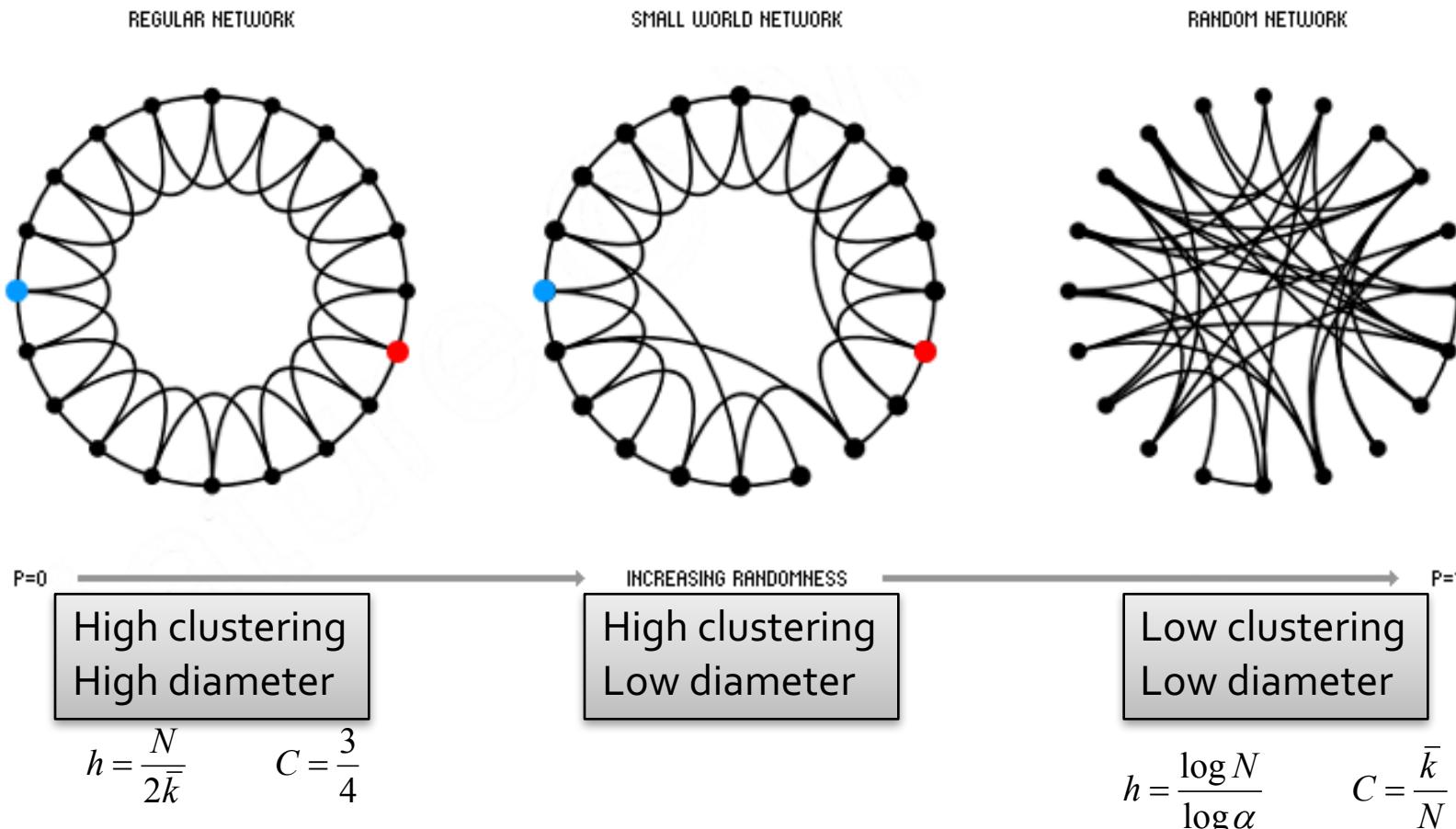
Small-World Model [Watts-Strogatz '98]

Two components to the model:

- **(1) Start with a low-dimensional regular lattice**
 - (In our case we are using a ring as a lattice)
 - Has high clustering coefficient
- **Now introduce randomness (“shortcuts”)**
- **(2) Rewire:**
 - Add/remove edges to create shortcuts to join remote parts of the lattice
 - For each edge with prob. p move the other end to a random node

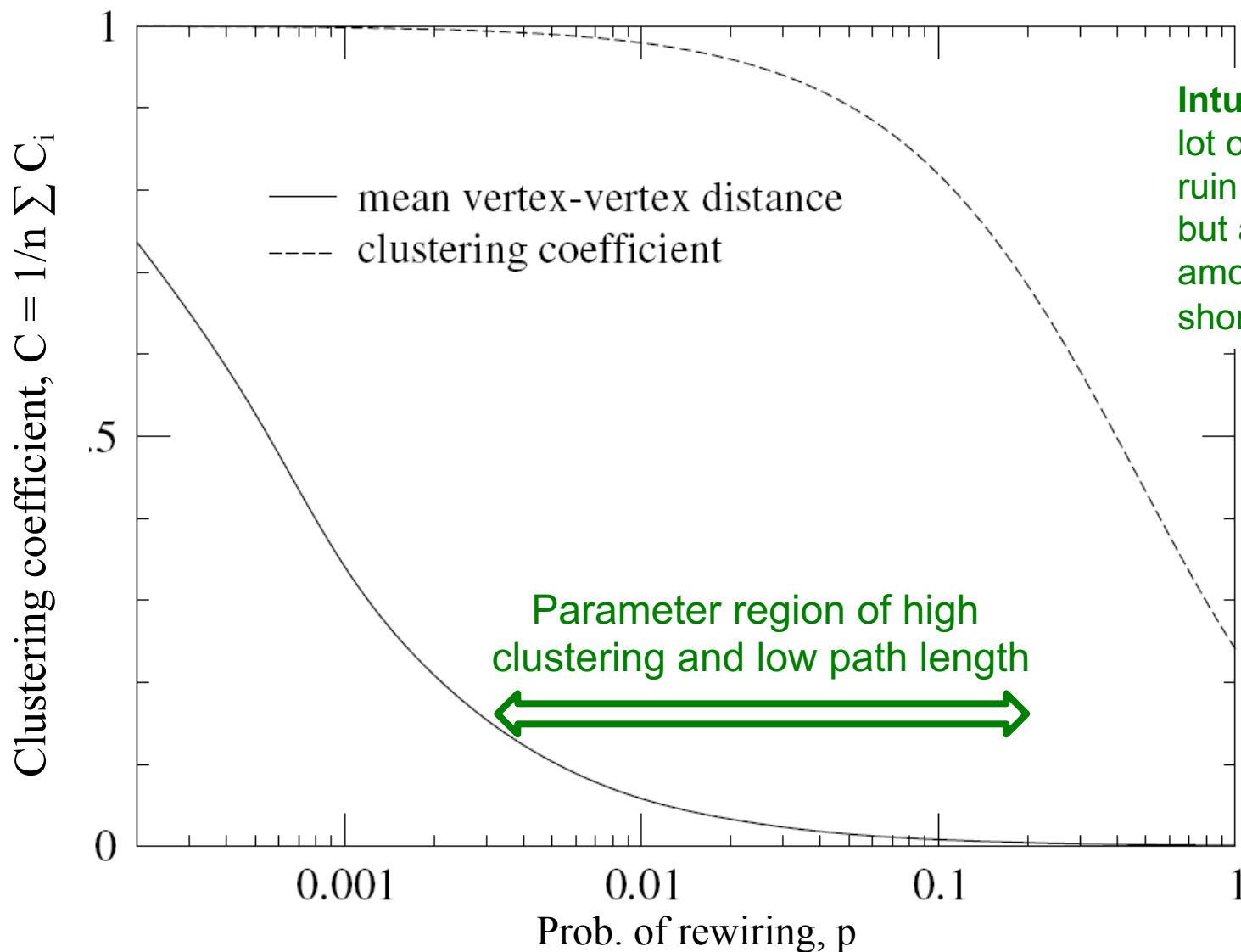


The Small-World Model



Rewiring allows us to “interpolate” between a regular lattice and a random graph

The Small-World Model

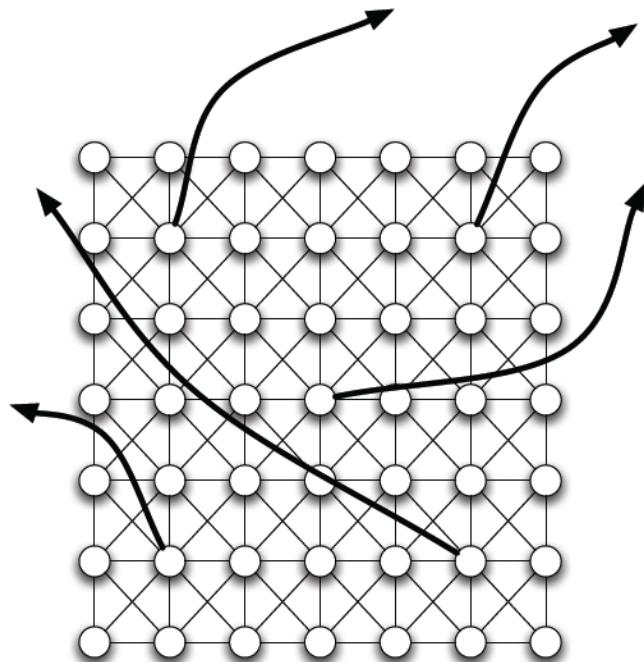


Intuition: It takes a lot of randomness to ruin the clustering, but a very small amount to create shortcuts.

Diameter of the Watts-Strogatz

■ Alternative formulation of the model:

- Start with a square grid
- Each node has 1 random long-range edge
 - Each node has 1 spoke. Then randomly connect them.



$$C_i = \frac{2 \cdot e_i}{k_i(k_i - 1)} = \frac{2 \cdot 12}{9 \cdot 8} \geq 0.33$$

There are already 12 triangles in the grid and the long-range edge can only close more.

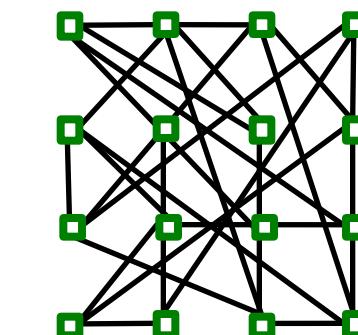
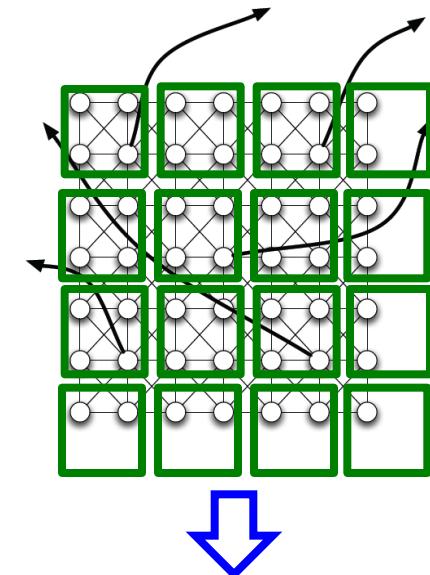
What's the diameter?
It is $O(\log(n))$
Why?

Diameter of the Watts-Strogatz

■ Proof:

- Consider a graph where we contract 2×2 subgraphs into supernodes
- Now we have 4 edges sticking out of each supernode
 - **4-regular random graph!**
- From Thm. we have short paths between super nodes
- We can turn this into a path in a real graph by adding at most 2 steps per long range edge (by having to traverse internal nodes)

⇒ **Diameter of the model is $O(2 \log n)$**



4-regular random graph

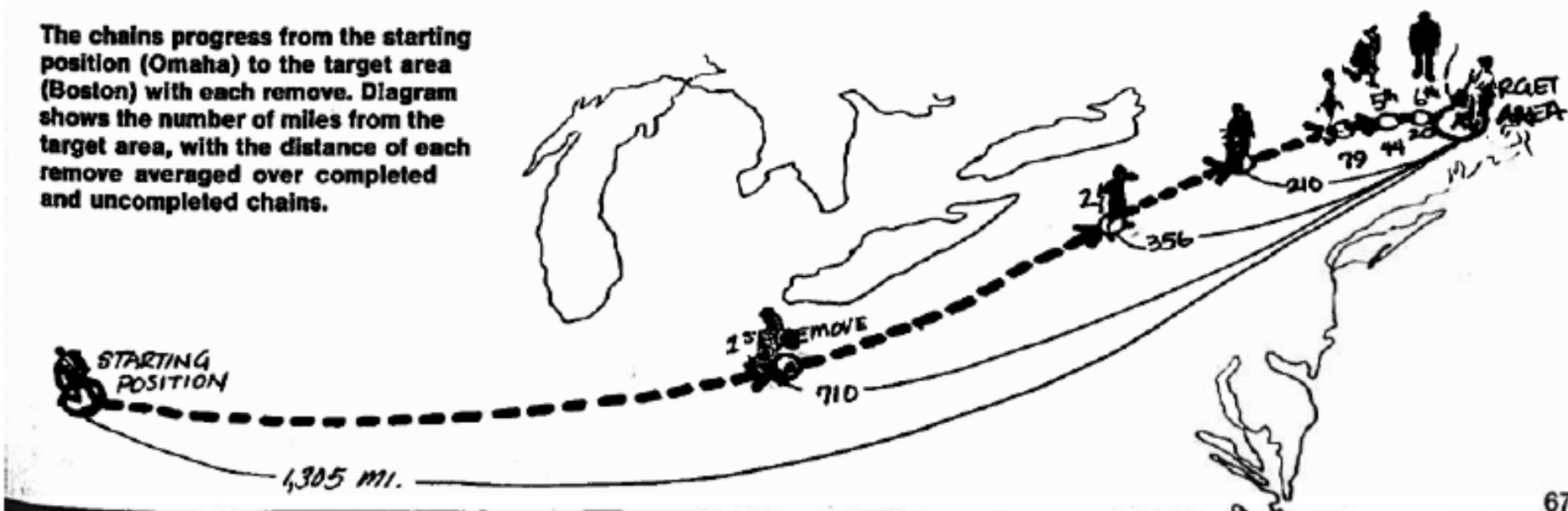
Small-World: Summary

- Could a network with high clustering be at the same time a small world?
 - Yes! You don't need more than a few random links
- The Watts Strogatz Model:
 - Provides insight on the interplay between clustering and the small-world
 - Captures the structure of many realistic networks
 - Accounts for the high clustering of real networks
 - Does not lead to the correct degree distribution
 - Does not enable navigation (next)

How to Navigate a Network?

- What mechanisms do people use to navigate networks and find the target?

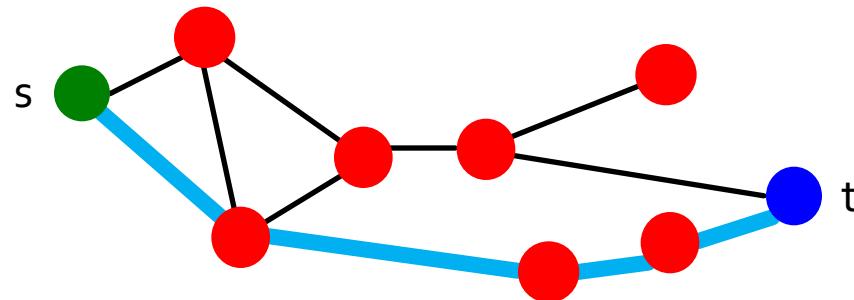
The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.



Decentralized Search

The setting:

- s only knows **locations** of its friends and location of the **target t**
- s does not know links of anyone else but itself
- **Geographic Navigation:** s “navigates” to a node geographically closest to t
- **Search time T:** Number of steps to reach t



Overview of the Results

Searchable

Search time T:

$$O((\log n)^\beta)$$

Kleinberg's model

$$O((\log n)^2)$$

Note: We know these graphs have diameter $O(\log n)$.
So in Kleinberg's model search time is polynomial in $\log n$,
while in Watts-Strogatz it is exponential (in $\log n$).

Not searchable

Search time T:

$$O(n^\alpha)$$

Watts-Strogatz

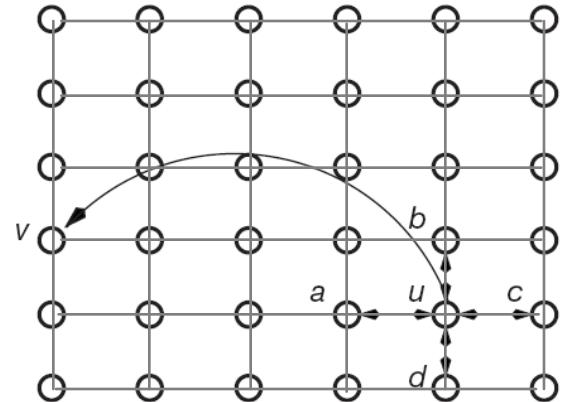
$$O(n^{\frac{2}{3}})$$

Erdős–Rényi

$$O(n)$$

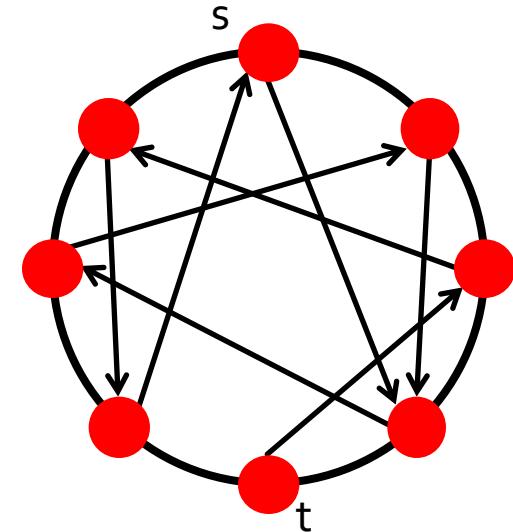
Navigation in Watts-Strogatz

- **Model:** 2-dim grid where each node has 1 random edge
 - This is a small-world!
 - (Small-world = diameter $O(\log n)$)
- **Fact:** A decentralized search algorithm in Watts-Strogatz model needs $n^{2/3}$ steps to reach t in expectation
 - Note: Even though paths of $O(\log n)$ steps exist
- Note: All our calculations are asymptotic, i.e., we are interested in what happens as $n \rightarrow \infty$



Navigation in Watts-Strogatz

- Let's do the proof for 1-dimensional case
- Want to show Watts-Strogatz is NOT searchable
 - Bound the search time from below
- About the proof:
 - Setting: n nodes on a ring plus one random directed edge per node.
 - Search time is $T \geq O(\sqrt{n})$
 - For d -dim. case: $T \geq O(n^{d/(d+1)})$
 - Proof strategy: Principle of deferred decision
 - Doesn't matter when a random decision is made if you haven't seen it yet
 - Assume random long range links are only created once you get to them



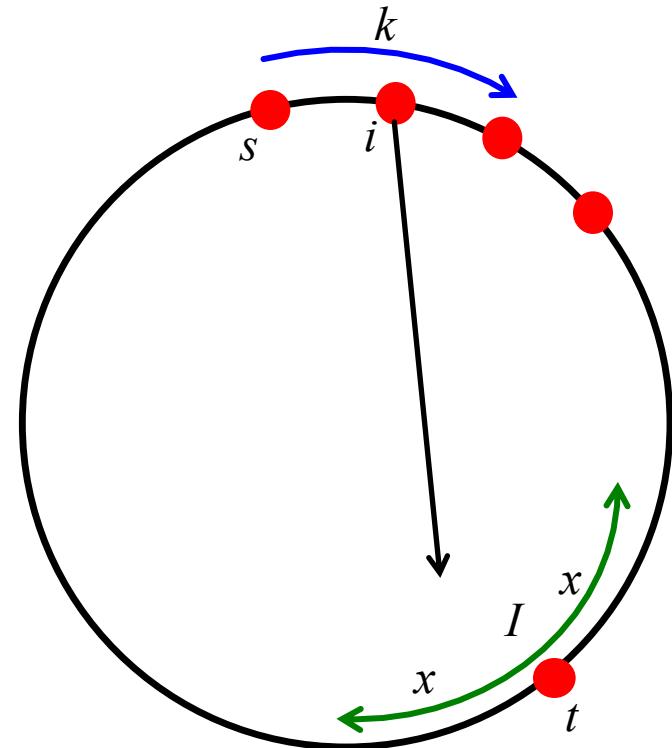
Proof: Search time is $\geq O(n^{1/2})$

- **Claim:**

- Expected search time is $\geq \frac{1}{4} \sqrt{n}$

- **Let:** E_i = event that long link out of node i points to some node in interval I of width $2 \cdot x$ nodes (for some x) around target t

- **Then:** $P(E_i) = \frac{2x}{n-1} \approx \frac{2x}{n}$ (in the limit of large n)
(haven't seen node i yet, but can assume random edge generation)



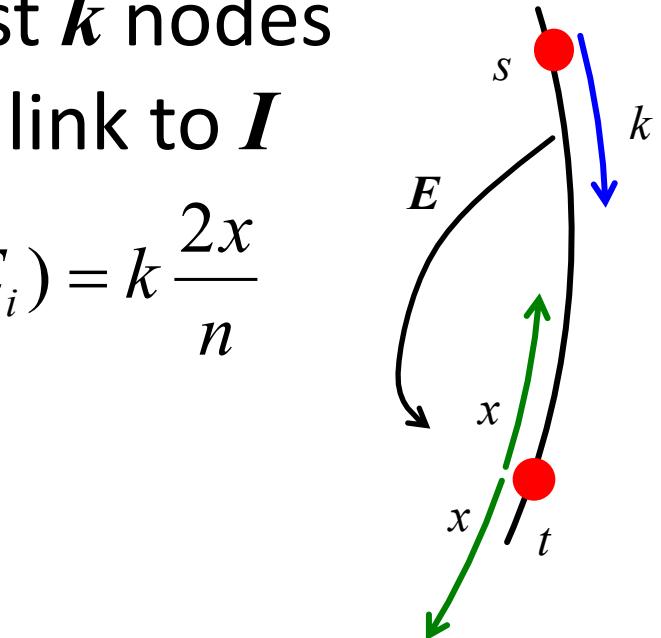
The basic intuition : The probability of jumping to somewhere close to t is small. It takes \sqrt{n} steps to jump into \sqrt{n} close to t with non-negligible probability.

Proof: Search time is $\geq O(n^{1/2})$

- E = event that any of the first k nodes search algorithm visits has a link to I
- **Then:** $P(E) = P\left(\bigcup_i^k E_i\right) \leq \sum_i^k P(E_i) = k \frac{2x}{n}$
- **Let's choose** $k = x = \frac{1}{2} \sqrt{n}$

Then:

$$P(E) \leq 2 \frac{\left(\frac{1}{2} \sqrt{n}\right)^2}{n} = \frac{1}{2}$$



Note: Our alg. is deterministic and will choose to travel via a long- or short-range links using some deterministic rule.

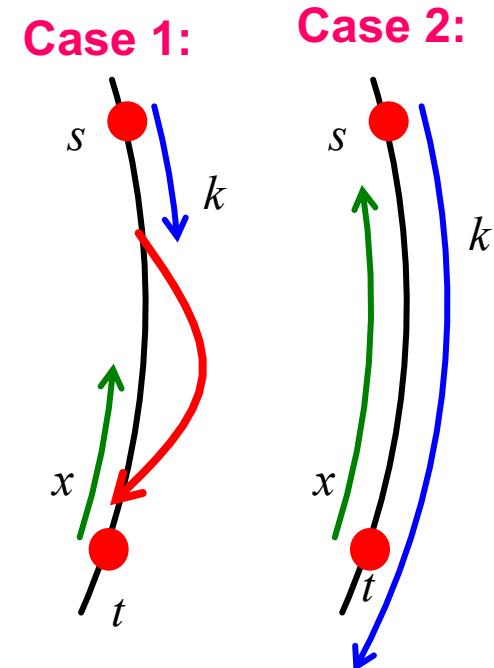
The principle of deferred decision tells us that it does not really matter how we reached node i .

Its prob. of linking to interval I is: $2x/n$.

Proof: Search time is $\geq O(n^{1/2})$

$$P(E) = P(\text{in } \frac{1}{2}\sqrt{n} \text{ steps we jump inside } \frac{1}{2}\sqrt{n} \text{ of } t) \leq \frac{1}{2}$$

- **Suppose** initial s is outside I and event E does not happen (i.e., first k visited nodes don't point to I)
- **Then** the search algorithm must take $T \geq \min(k, x)$ steps to get to t
 - (1) Right after we visit k nodes a good long-range link may occur
 - (2) x and k “overlap”, due to E not happening we have to walk at least x steps



Proof: Search time is $\geq O(n^{1/2})$

- **Claim:** Getting from s to t takes $\geq \frac{1}{4}\sqrt{n}$ steps
- $\text{Search time} \geq P(E)(\#\text{steps}) + P(\text{not } E) \min(x, k)$
- **Proof:** We just need to put together the facts

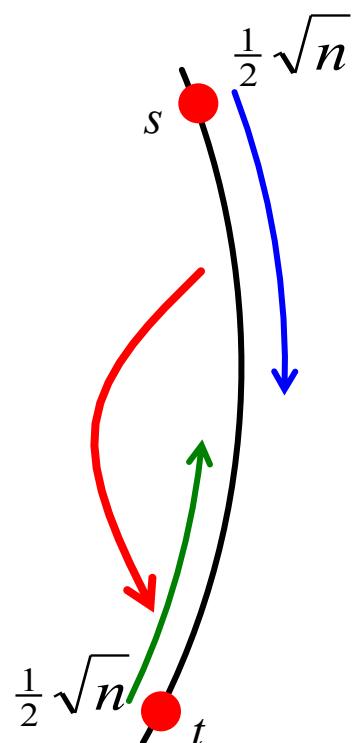
- We already showed that for $x = k = \frac{1}{2}\sqrt{n}$

- E does not happen with prob. $\frac{1}{2}$
- If E does not happen,
we must traverse $\geq \frac{1}{2}\sqrt{n}$ steps to get to t

- The expected time to get to t is then

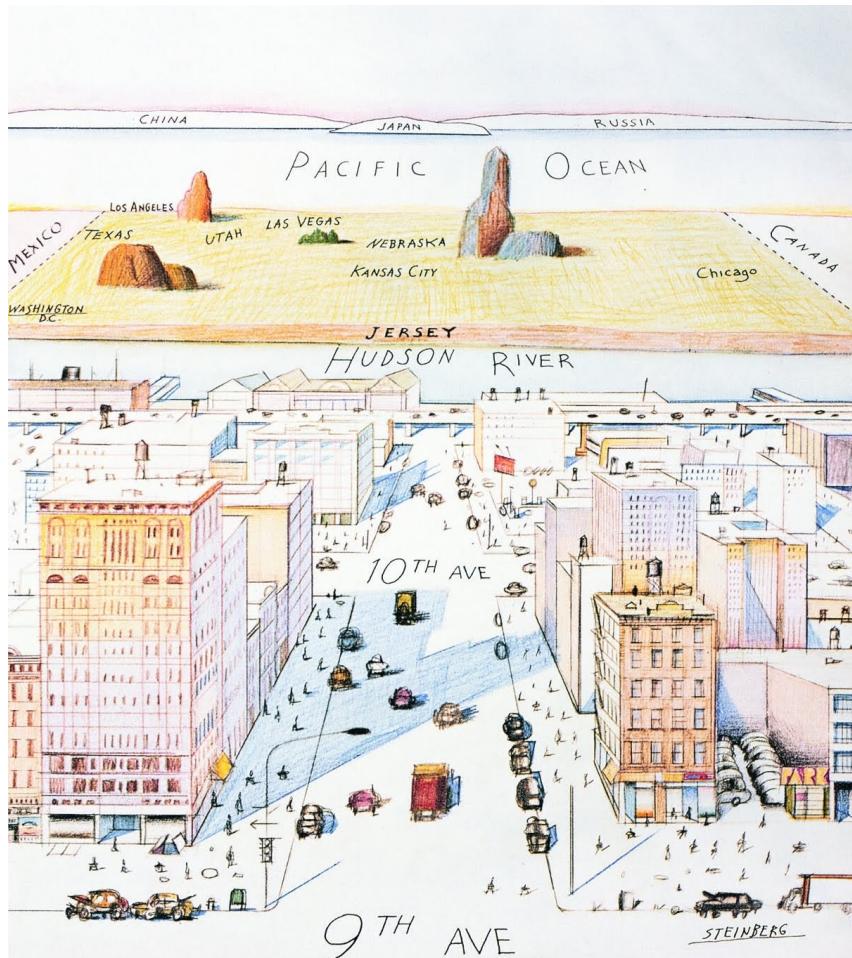
$$\geq P(E \text{ doesn't occur}) \cdot \min\{x, k\} =$$

$$= \frac{1}{2} \frac{1}{2} \sqrt{n} = \frac{1}{4} \sqrt{n}$$



Navigable Small-World Graph?

- Watts-Strogatz graphs are **not searchable**
- **How do we make a searchable small-world graph?**
- **Intuition:**
 - Our long range links are not random
 - **They follow geography!**



Saul Steinberg, "View of the World from 9th Avenue"