Small World Networks

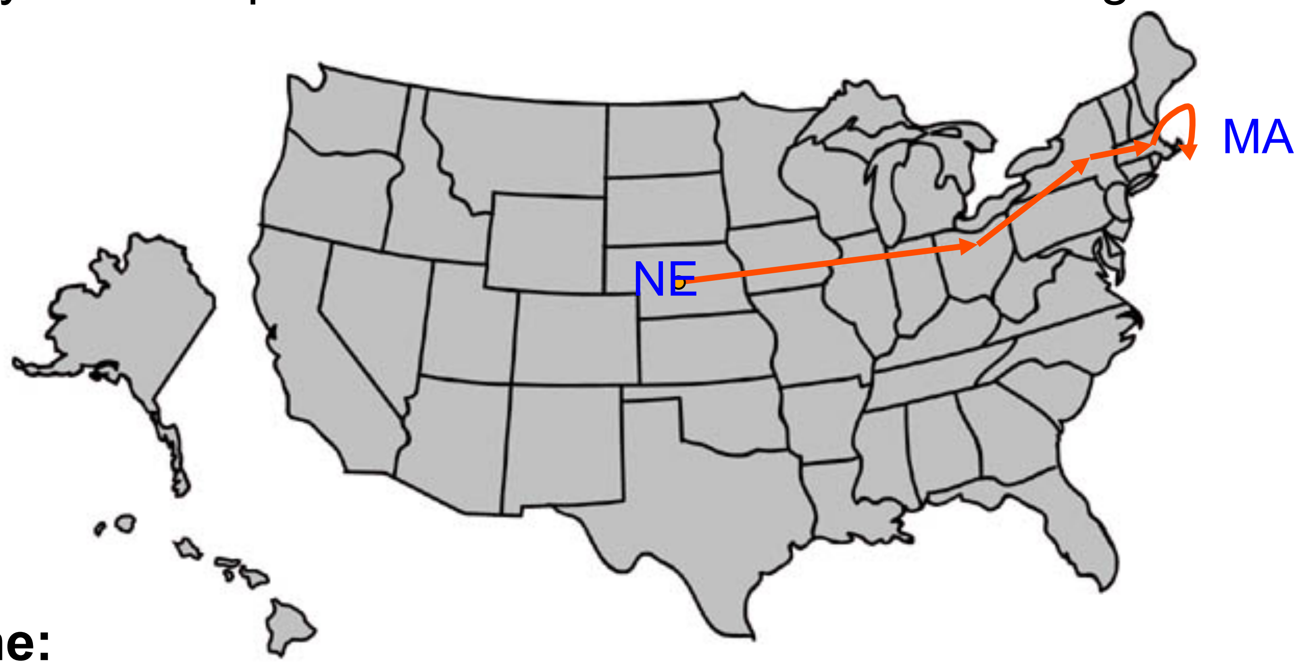
Adapted from slides by Lada Adamic, UMichigan

Outline

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

Small World Phenomenon: 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”**

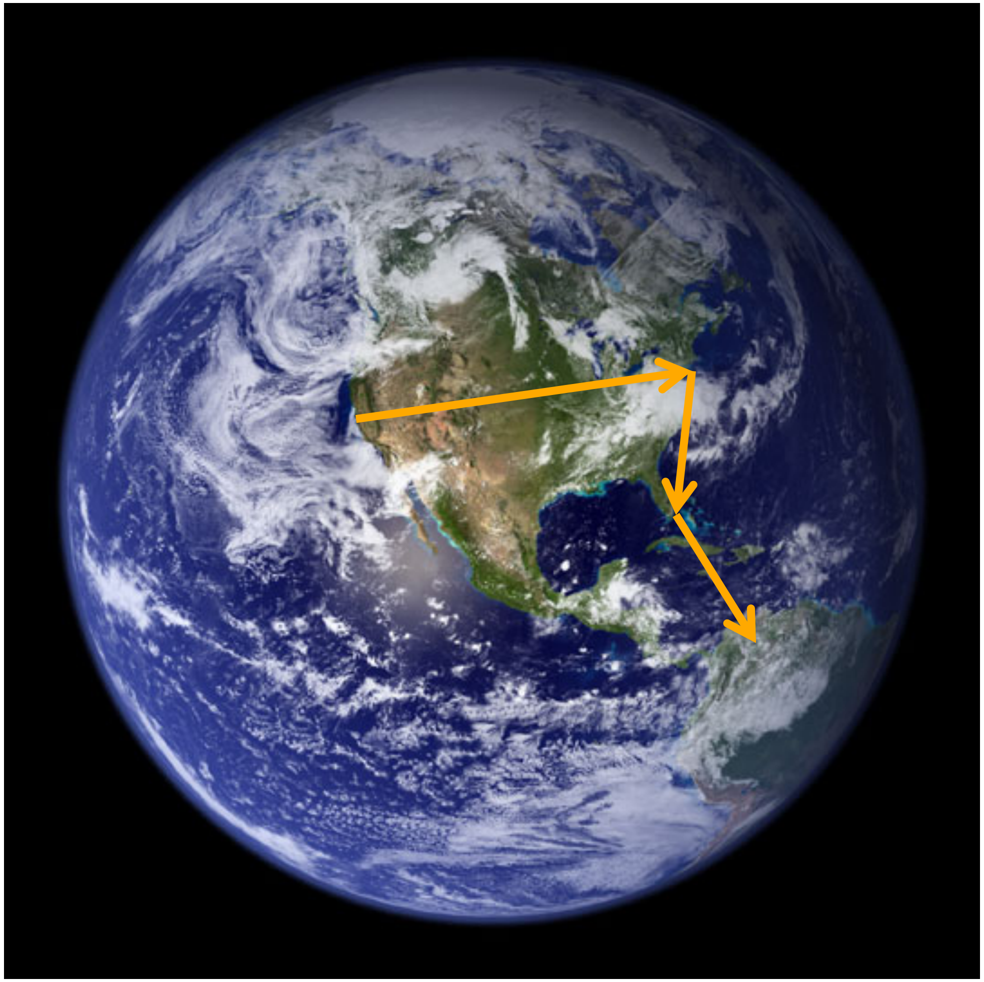
Source: undetermined

Small World Phenomenon:
Milgram's Experiment Repeated

email experiment by Dodds, Muhamad, Watts; Science 301, (2003) (reading linked on website)

- 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

Small World Phenomenon: Interpreting Milgram's experiment

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

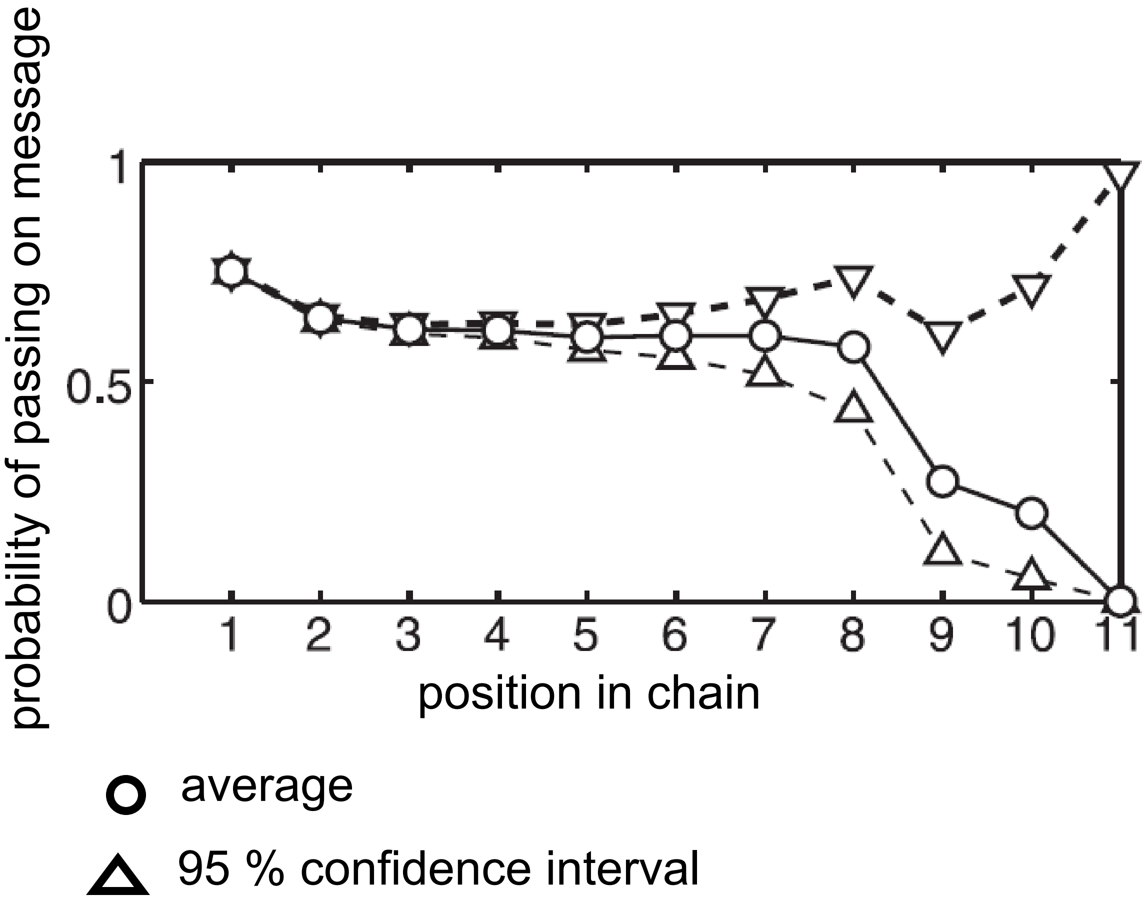
- If social networks were random… ?
  - Pool and Kochen (1978) - ~500-1500 acquaintances/person
  - ~ 1,000 choices 1st link
  - ~ 1000^2 = 1,000,000 potential 2nd links
  - ~ 1000^3 = 1,000,000,000 potential 3rd links

- If networks are completely cliquish:
  - all my friends' friends are my friends
  - What would happen?

Small world experiment: Accuracy of distances

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

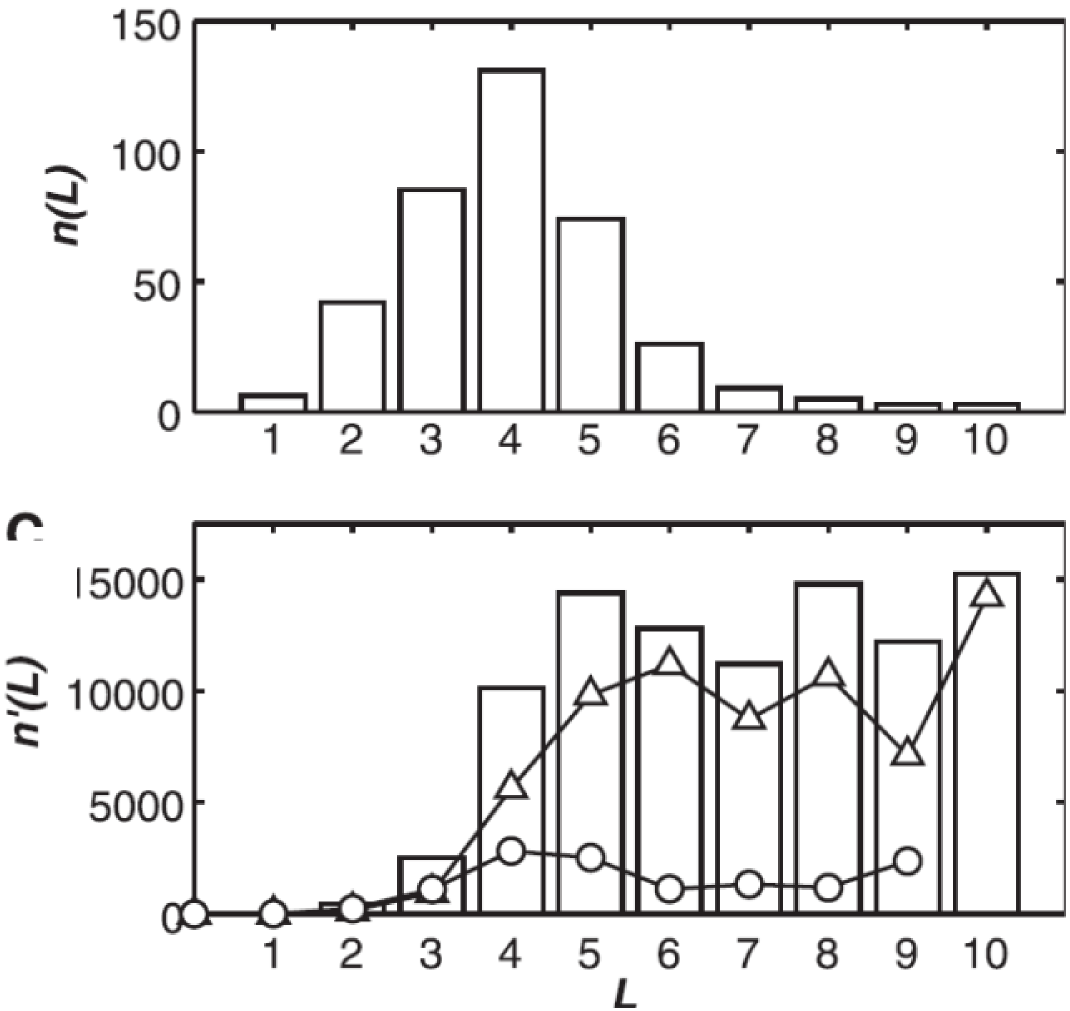
Small world experiment accuracy:
Attrition rate is approx. constant



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.

Small world experiment accuracy:

Estimating true distance distribution



- observed chain lengths
- 'recovered' histogram of path lengths

○ 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.

Small world experiment:
Accuracy of distances

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

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

Current Social Networks

- Facebook's data team released two papers in Nov. 2011
  - 721 million users with 69 billion friendship links
  - Average distance of 4.74
- Twitter studies
  - Sysomos reports the average distance is 4.67 (2010)
    - 50% of people are 4 steps apart, nearly everyone is 5 steps or less
  - Bakhshandeh et al. (2011) report an average distance of 3.435 among 1,500 random Twitter users

Small world phenomenon:
Business applications?

“Social Networking” as a Business:

- Facebook, Google+, Orkut, Friendster
  entertainment, keeping and finding friends

- LinkedIn:
  - more traditional networking for jobs

- Spoke, VisiblePath
  - helping businesses capitalize on existing client relationships

Small world phenomenon:
Applicable to other kinds of networks

Same pattern:

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

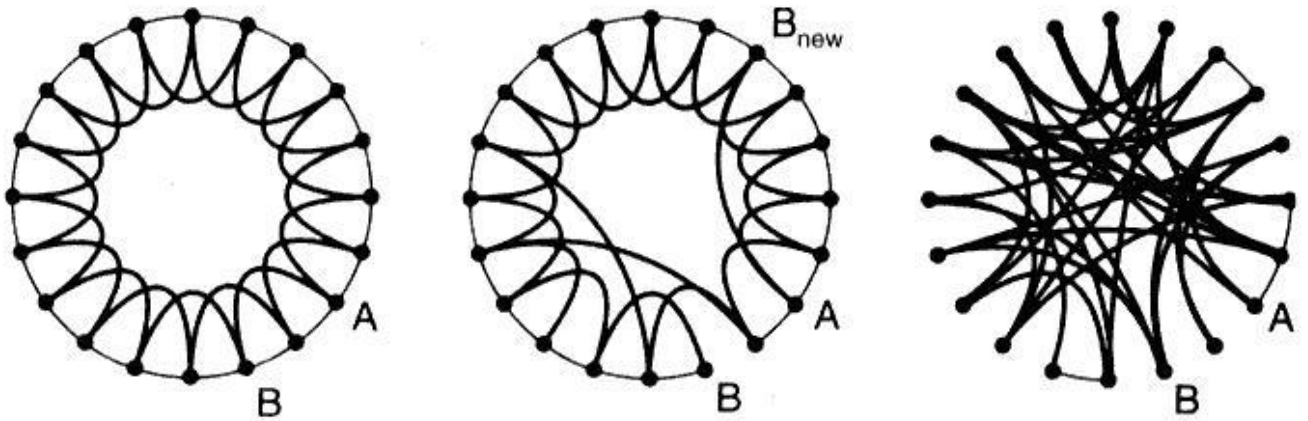
low average shortest path $l_{\text{network}} \approx \ln(N)$

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph
- food webs

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 one of their endpoints

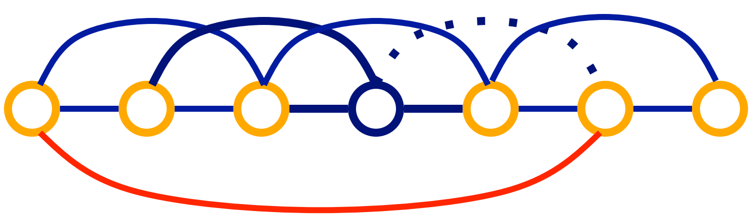
Add a fraction p of additional edges leaving underlying lattice intact

- 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



Watts/Strogatz model: What happens in between?

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

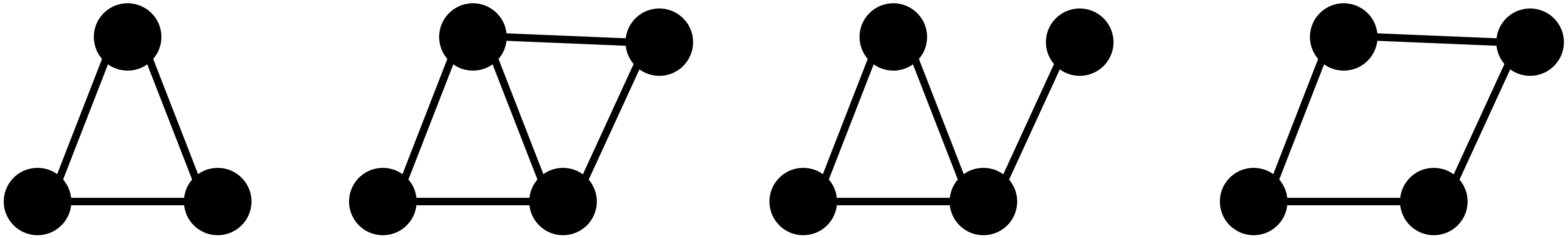
Clustering Coefficient

- Clustering coefficient for graph:

$$\frac{\#\ \text{triangles x 3}}{\#\ \text{connected triples}}$$

Each triangle gets counted 3 times

  - Also known as the “fraction of transitive triples”



Localized Clustering Coefficient

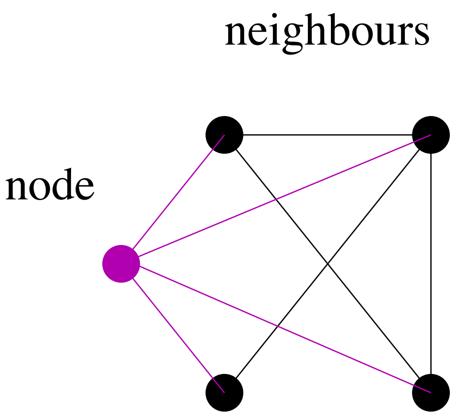
- Clustering for node v:

$$\frac{\text{\# actual edges between neighbors of v}}{\text{\# possible edges between neighbors of v}}$$

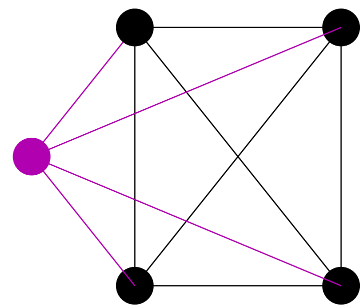
- Number of possible edges between k vertices: k(k-1)/2
  - i.e., the number of edges in a complete graph with k vertices
- Clustering coefficient for a vertex v with k neighbors

$$C(v) = \frac{|\text{actual edges}|}{k(k-1)/2} = \frac{2 \times |\text{actual edges}|}{k(k-1)}$$

Localized Clustering Coefficient



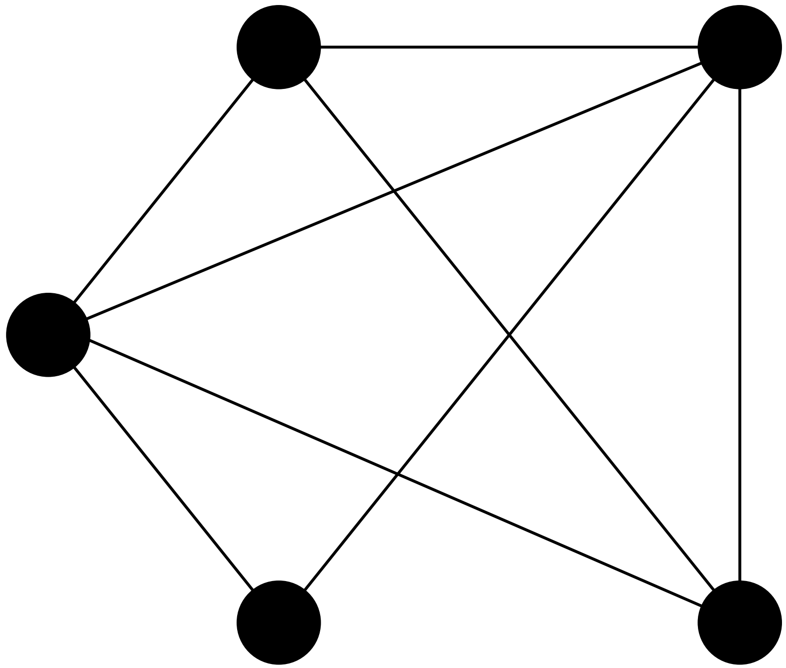
4 actual edges



$$\frac{4 \text{ x } 3}{2} = 6 \text{ possible edges}$$

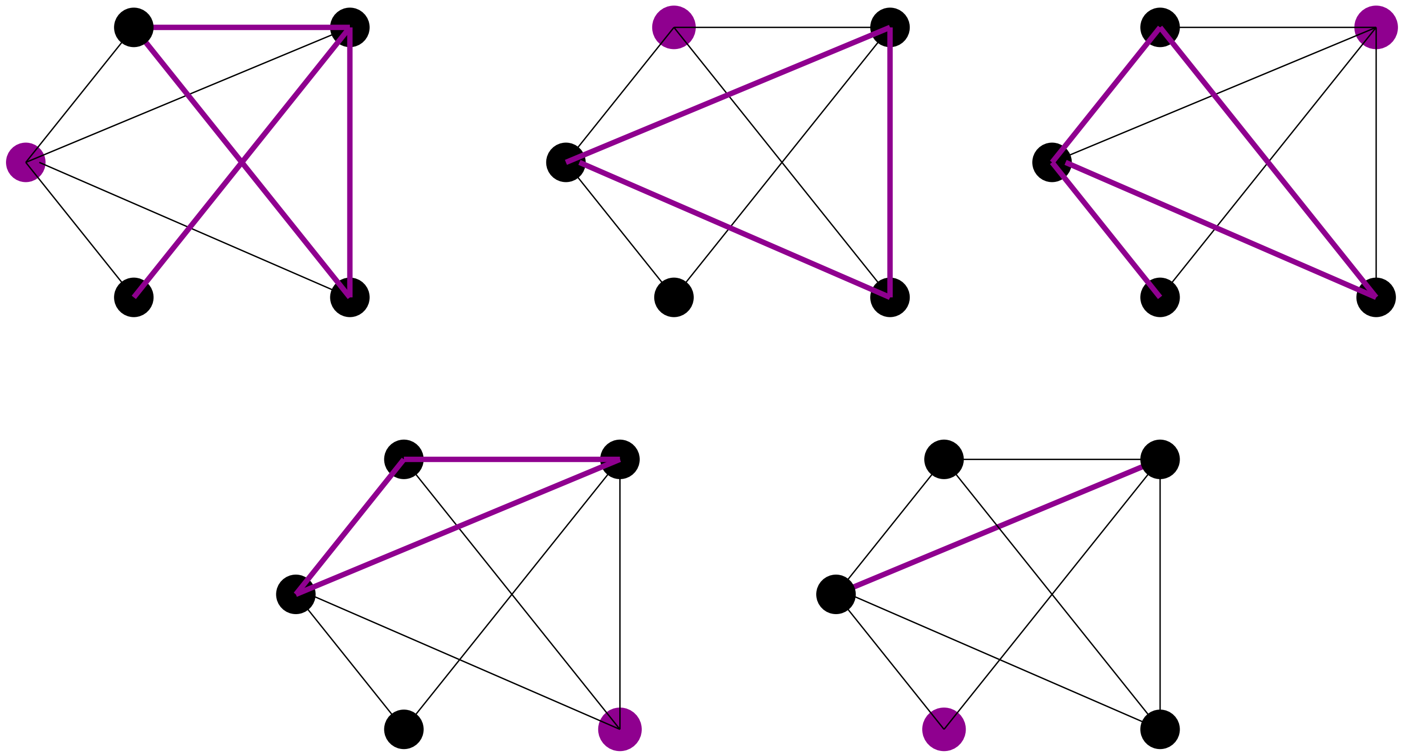
$$\text{Clustering : } \frac{4}{6} = 0.66$$

What is the average localized clustering coefficient?



Slide by Uta Priss (Edinburgh Napier U)

What is the average localized clustering coefficient?

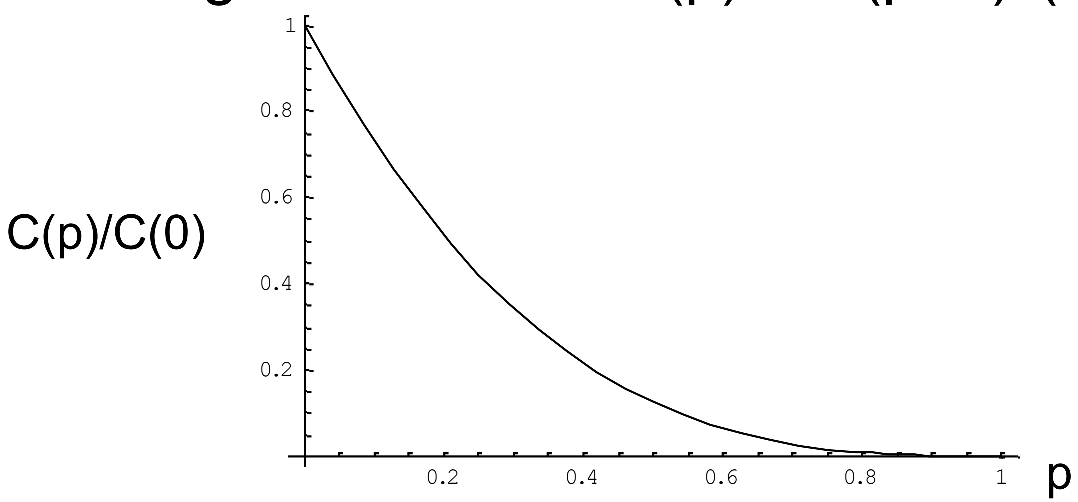


Slide by Uta Priss (Edinburgh Napier U)

Watts/Strogatz model: Clustering coefficient can be computed for SW model with rewiring

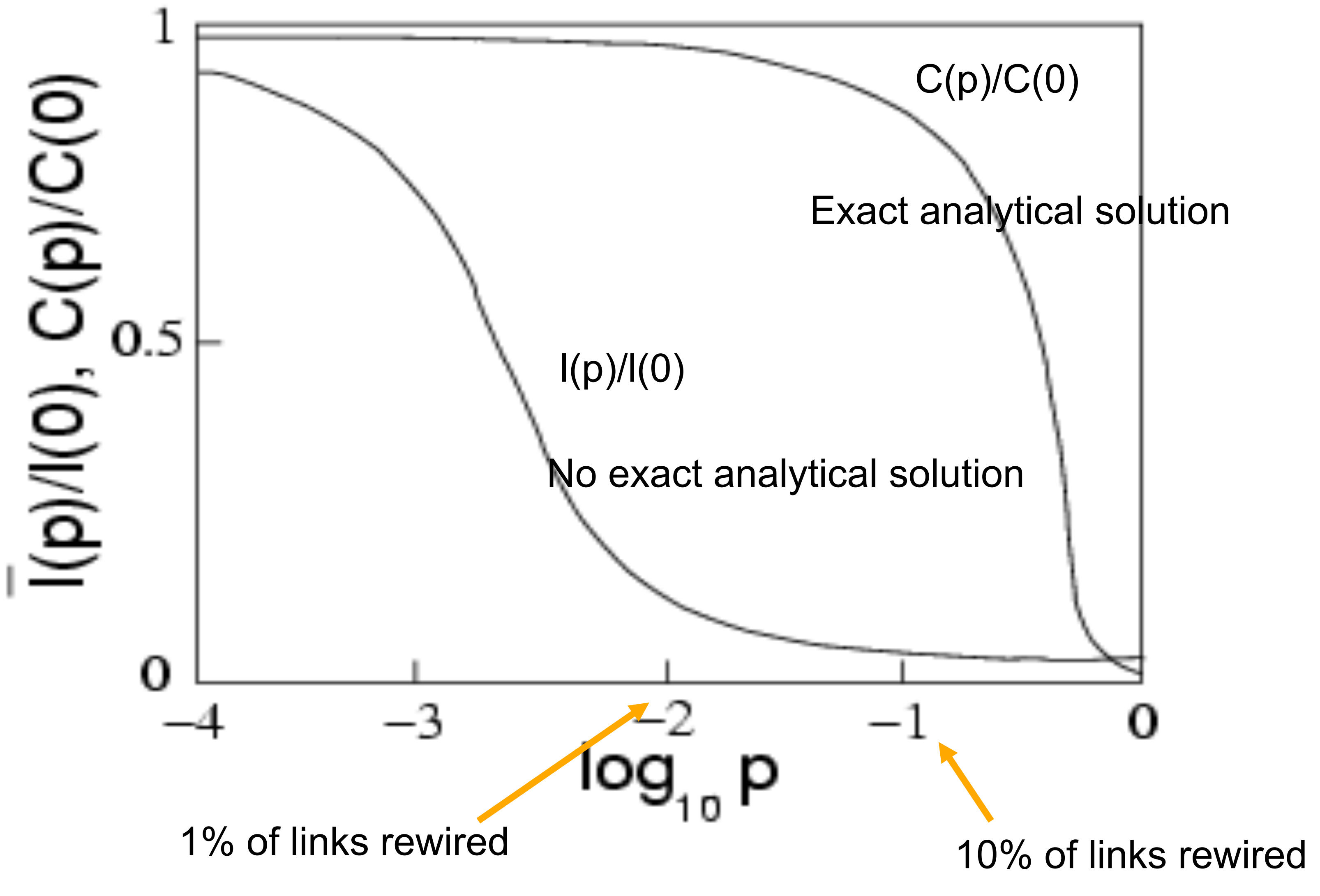
- The probability that a connected triple stays connected after rewiring
  - probability that none of the 3 edges were rewired $(1-p)^3$
  - probability that edges were rewired back to each other very small, can ignore

- 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.

Watts/Strogatz model: Change in clustering coefficient and average path length



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

Small-World Networks and Clustering

- A graph G is considered small-world, if:
  - its average clustering coefficient $\overline{C}_G$ is significantly higher than the average clustering coefficient of a random graph $\overline{C}_{rand}$ constructed on the same vertex set, and
  - the graph has approximately the same mean-shortest path length L_{sw} as its corresponding random graph L_{rand}

$$\overline{C}_G >> \overline{C}_{rand} \qquad L_G \cong L_{rand}$$

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

What features of real social networks are missing from the small world model?

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

Small world networks: Summary

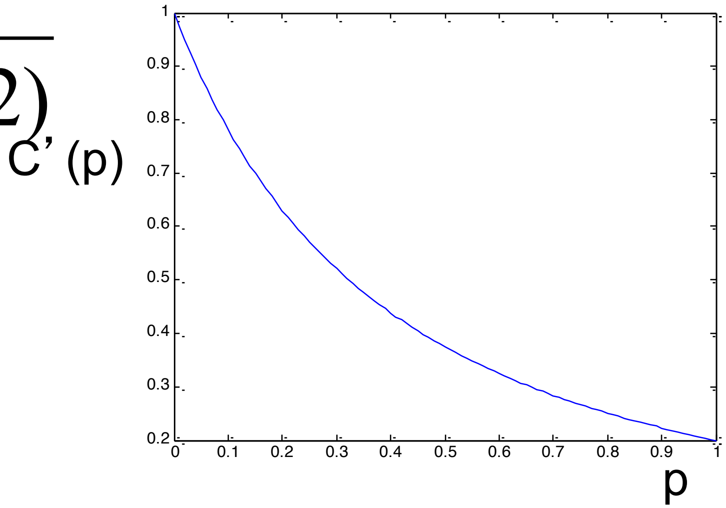
- The world is small!
- Watts & Strogatz came up with a simple model to explain why
- Other models incorporate geography and hierarchical social structure

Extra Material
(Not covered in class)

Watts/Strogatz model:
Clustering coefficient: addition of random edges

- How does C depend on p?
- C'(p)= 3xnumber of triangles / number of connected triples
- C'(p) computed analytically for the small world model without rewiring

$$C'(p) = \frac{3(k-1)}{2(2k-1)+4kp(p+2)}$$



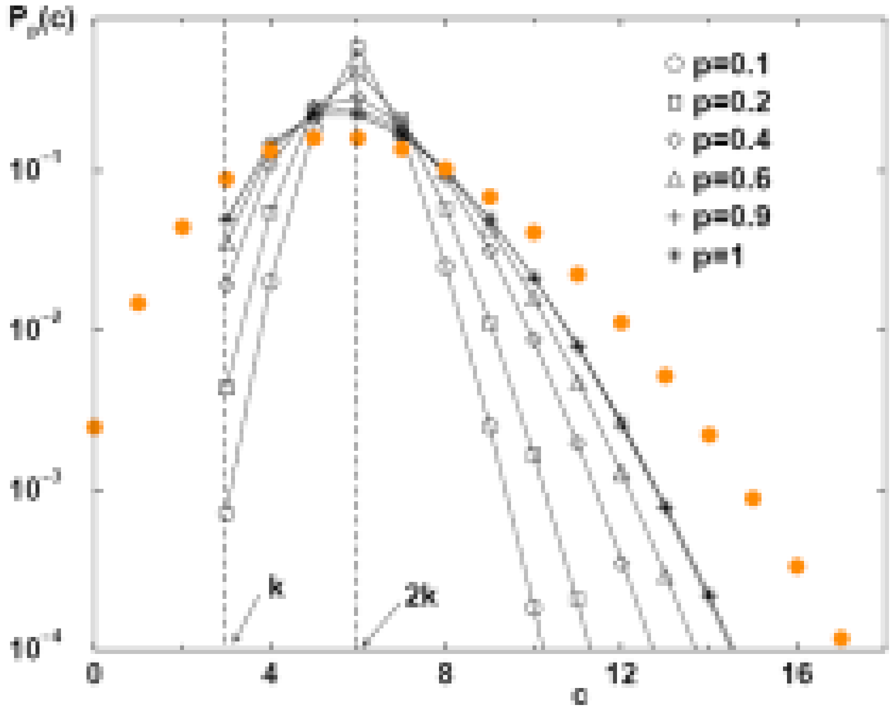
Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Watts/Strogatz model: Degree distribution

- p=0 delta-function
- p>0 broadens the distribution
- Edges left in place with probability (1-p)
- Edges rewired towards i with probability 1/N

Watts/Strogatz model:
Model: small world with probability p of rewiring

Even at p = 1, graph is not a purely random graph



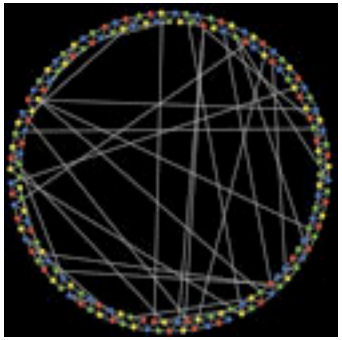
1000 vertices

● random network with average connectivity K

visit nodes sequentially and rewire links
exponential decay, all nodes have similar number of links

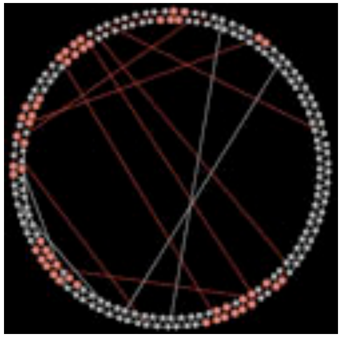
Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

demos: measurements on the WS small world graph



http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldWS.html

later on: see the effect of the small world topology on diffusion:

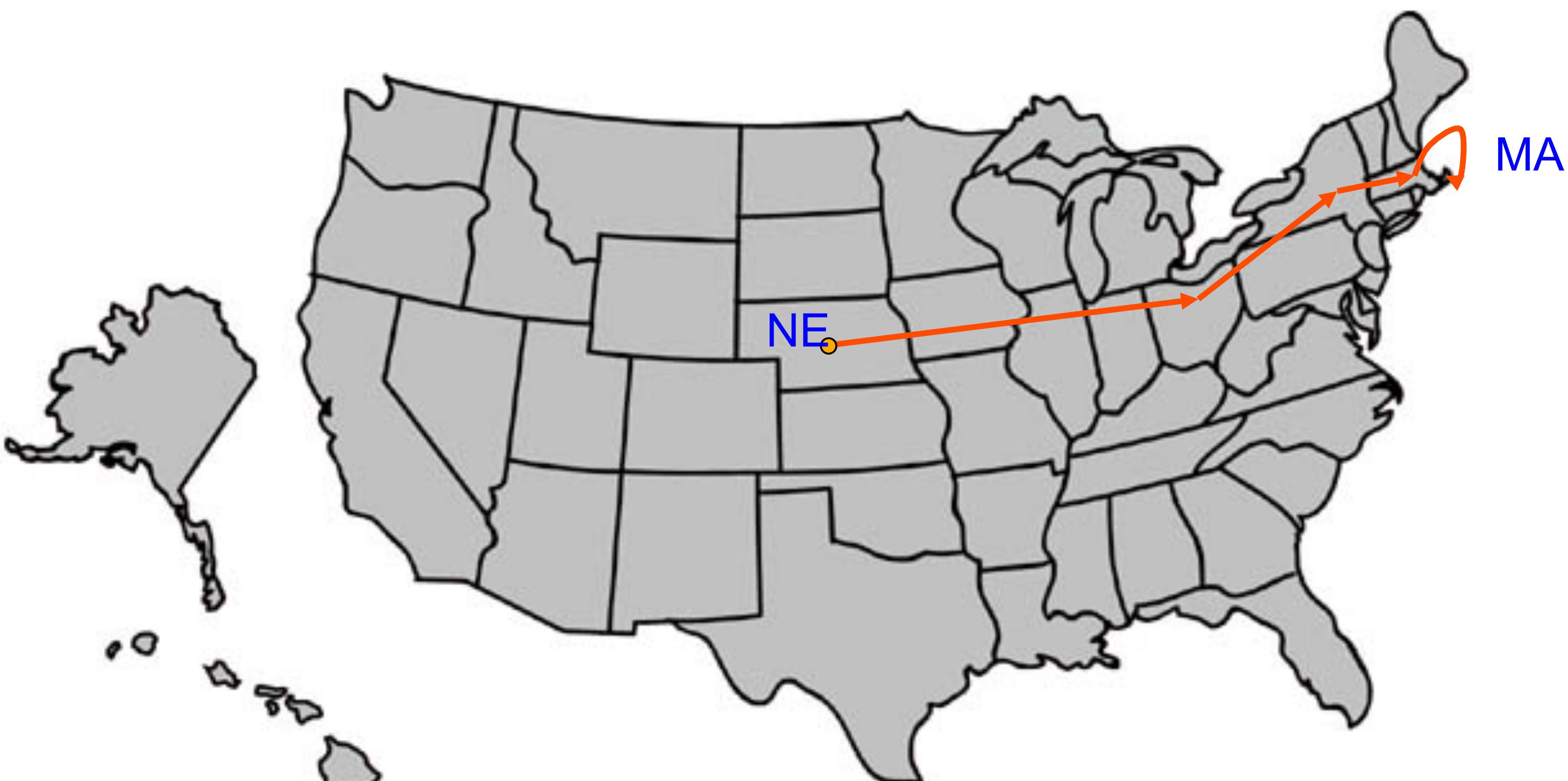


http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldDiffusionSIS.html

Geographical small world models: What if long range links depend on distance?

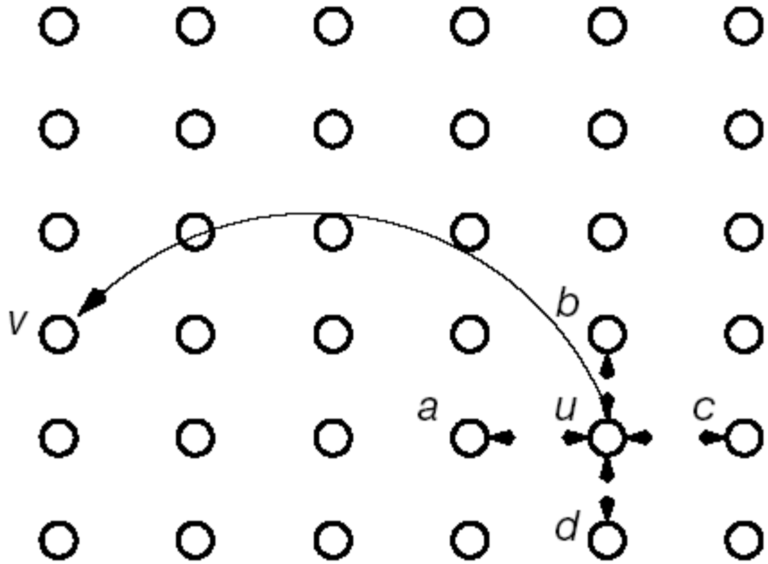
"The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain"

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



Source: undetermined

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

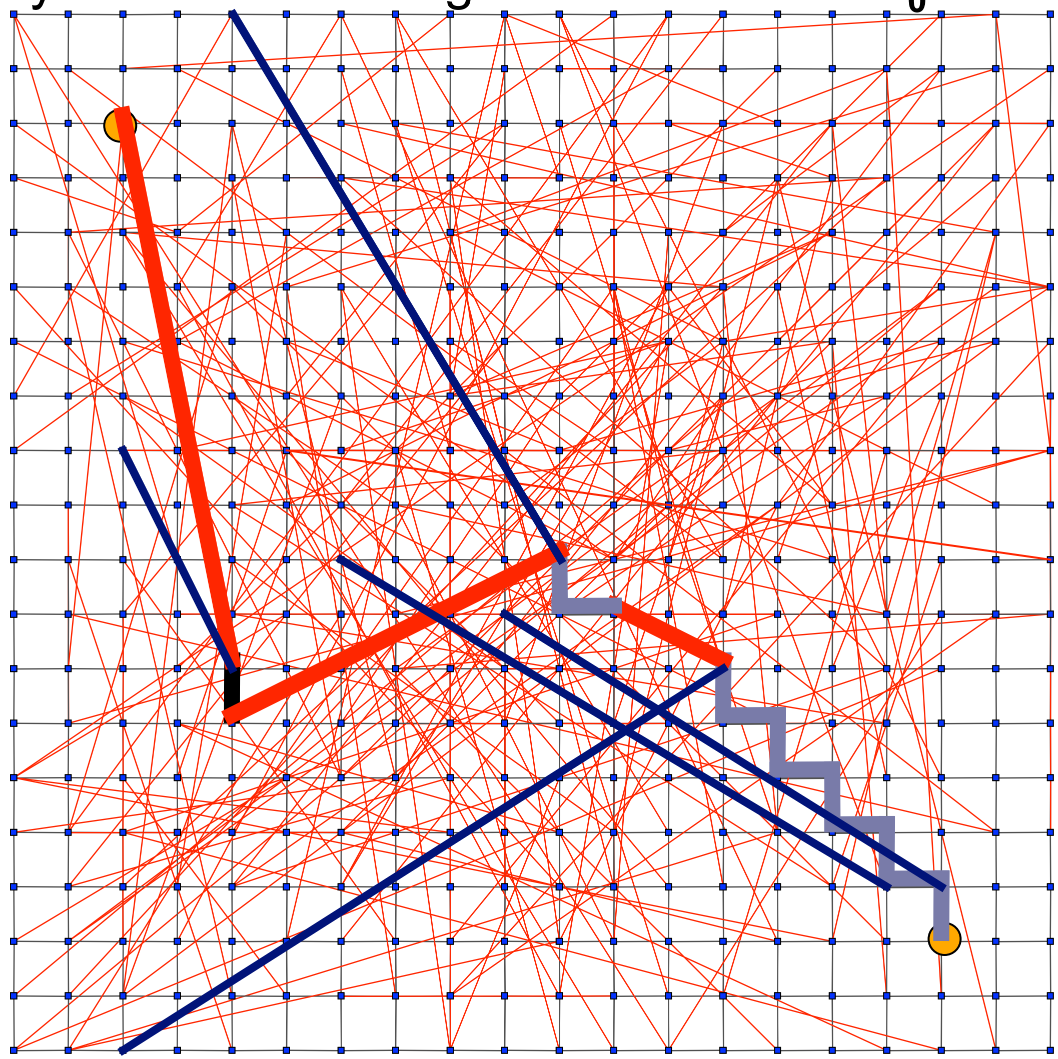
$$p(\text{link between u and v}) = (\text{distance}(u,v))^{-r}$$

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

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 $\mathbf{a_0 n^{2/3}}$



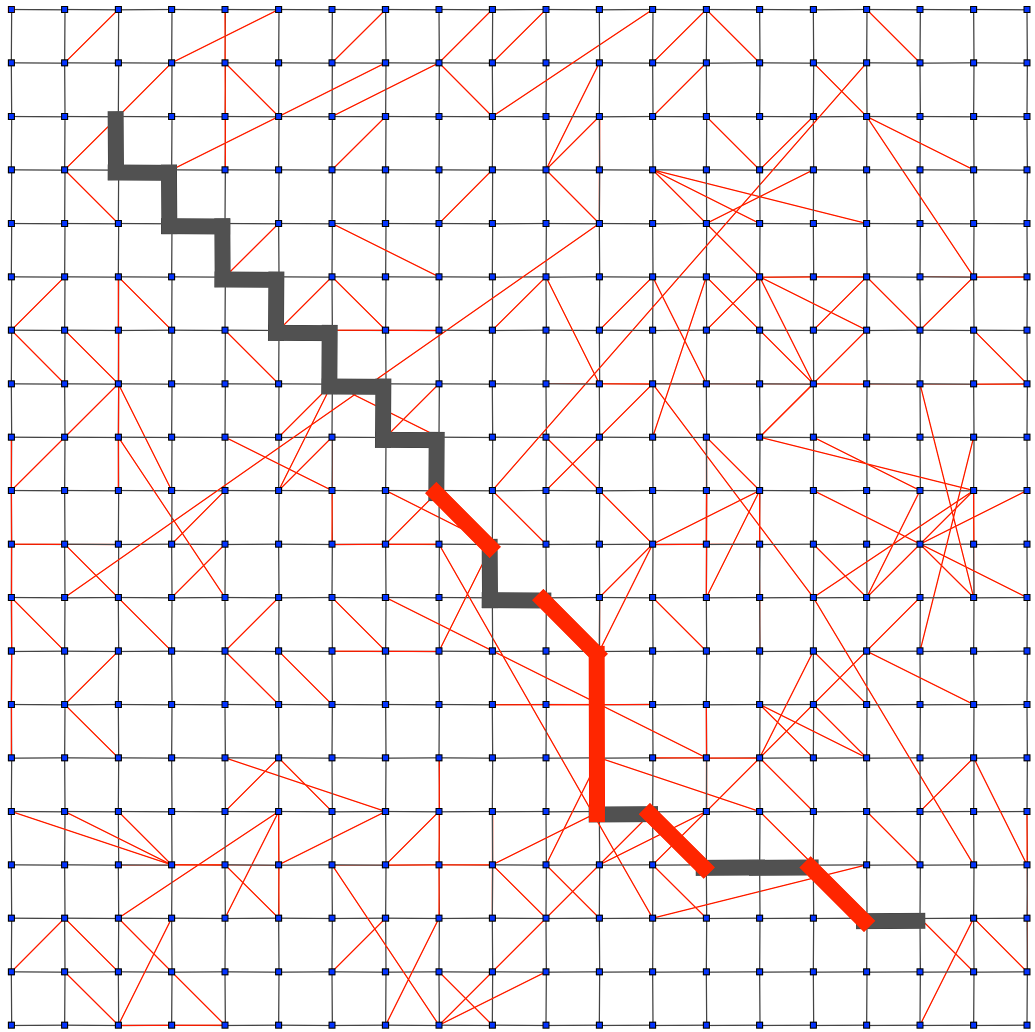
$p \sim p_0$

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 ~ $\mathbf{N}^{(r-2)/(r-1)}$

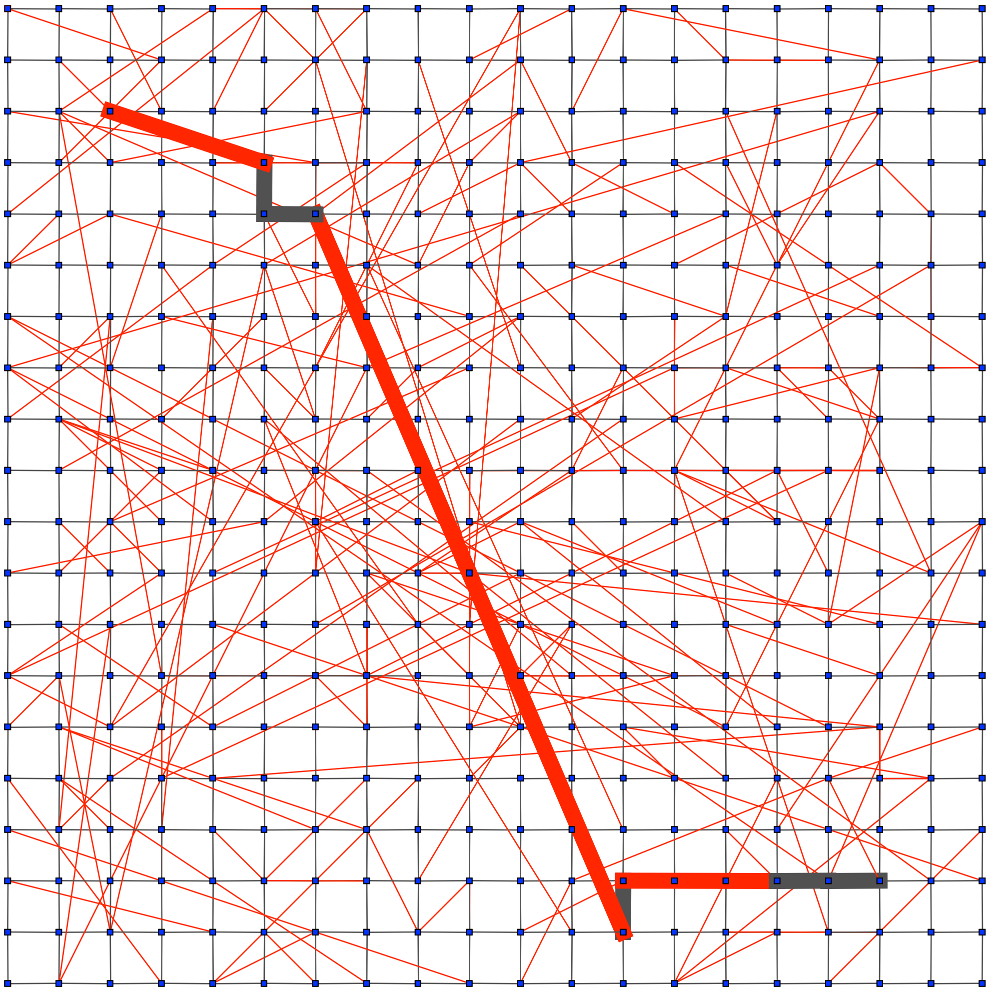
$$p \sim \frac{1}{d^4}$$



geographical small world model
Links balanced between long and short range

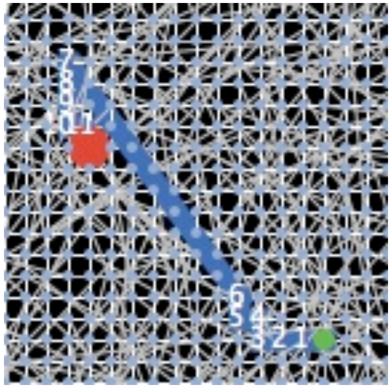
When **r=2,** expected time of a DA is at most $\mathbf{C\ (log\ N)^2}$

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



demo (a few weeks from now)

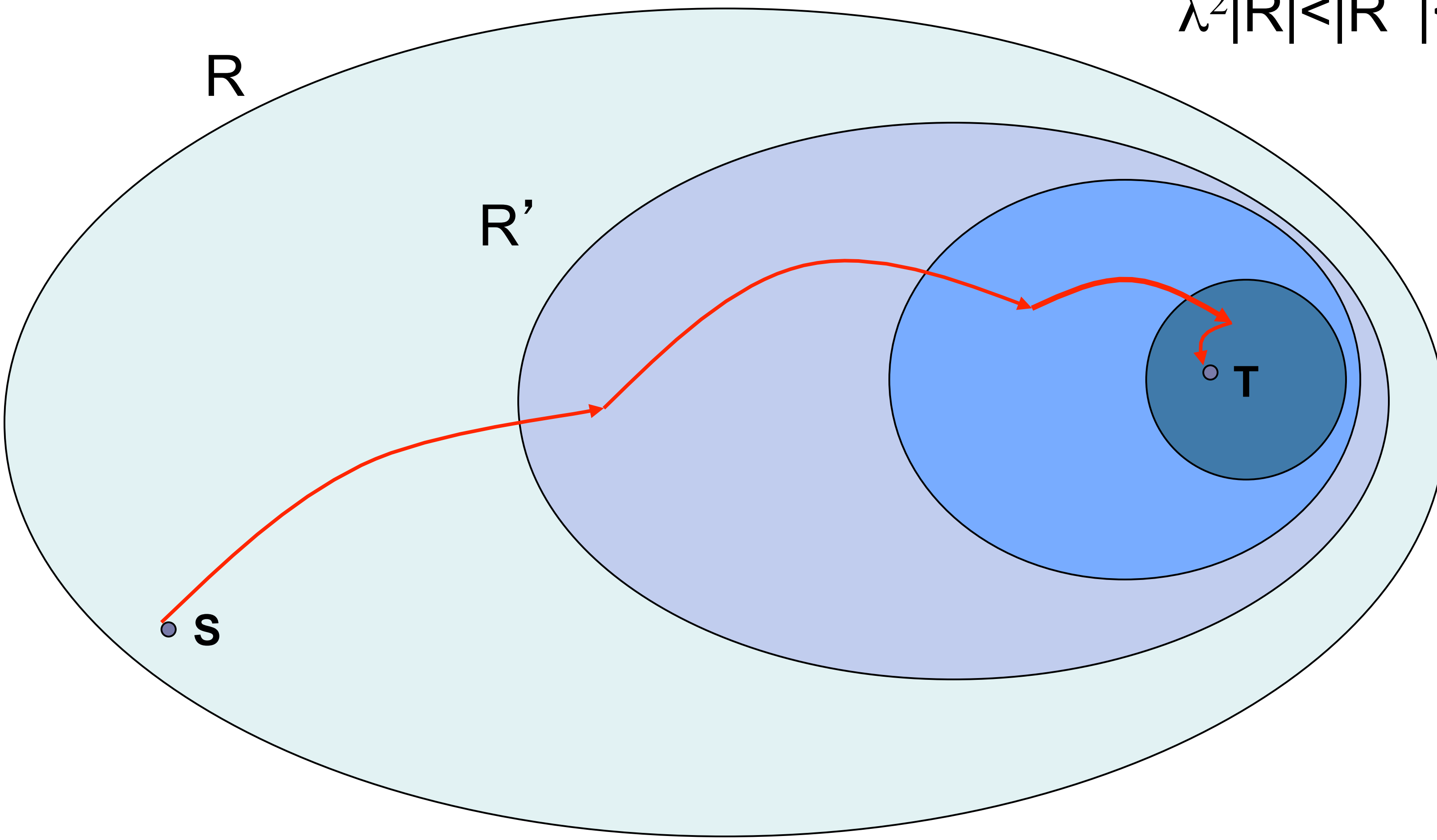
- how does the probability of long-range links affect search?



http://projects.si.umich.edu/netlearn/NetLogo4/SmallWorldSearch.html

Geographical small world model: navigability

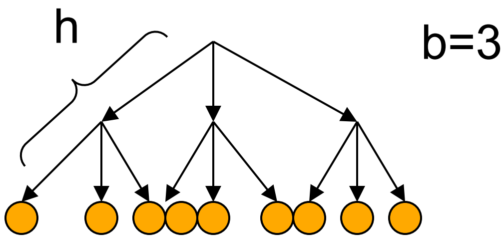
$\lambda^2|R|<|R'|<\lambda|R|$



$k = c \log^2 n$

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

hierarchical small-world models: Kleinberg



Hierarchical network models:

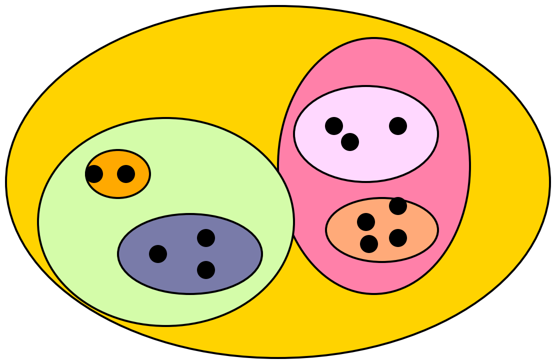
Individuals classified into a hierarchy,
h_{ij} = height of the least common ancestor.

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

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

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

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

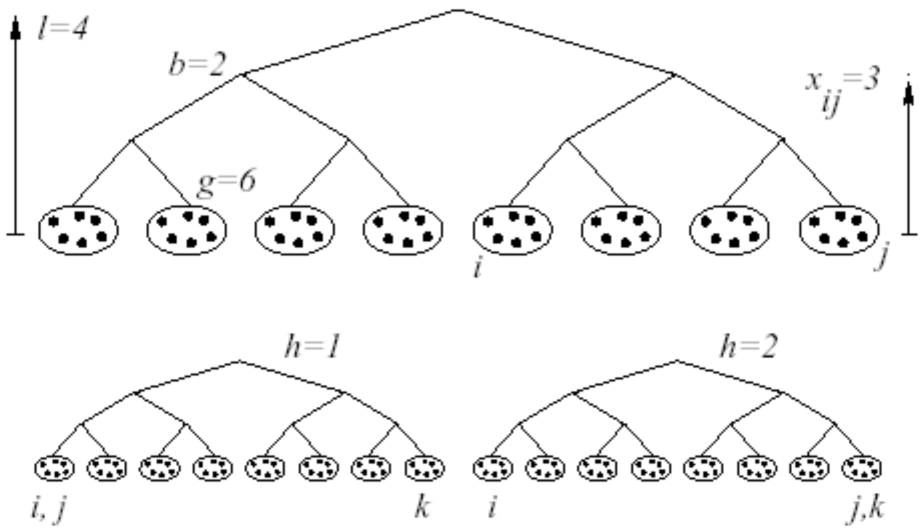


Source: Kleinberg, 'Small-World Phenomena and the Dynamics of Information' NIPS 14, 2001.

Hierarchical small world models:

Watts, Dodds, Newman (Science, 2001)

individuals belong to hierarchically nested groups

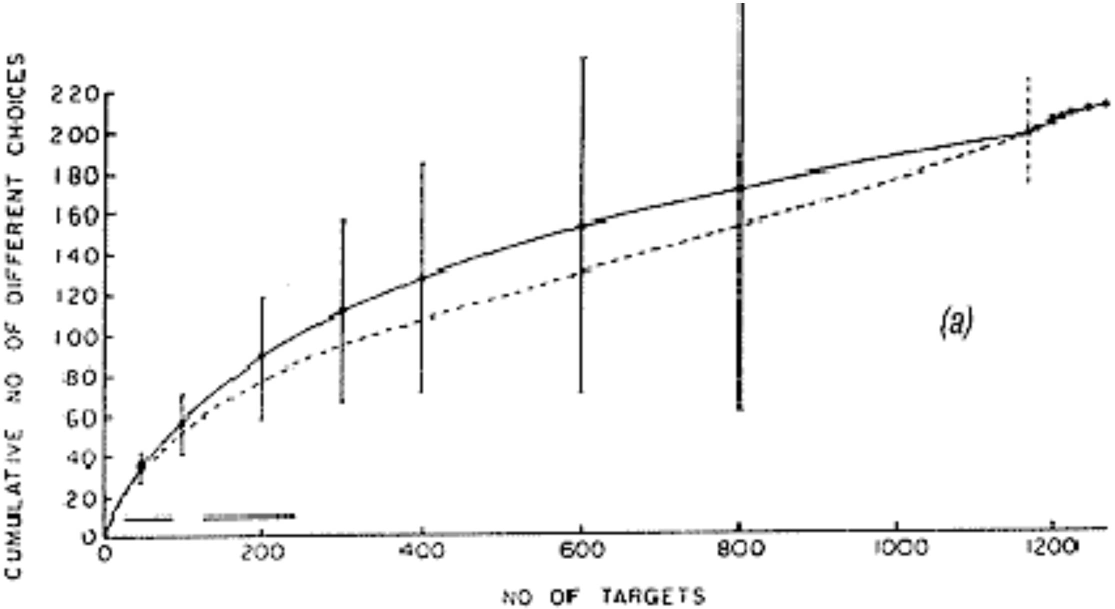


$p_{ij} \sim \exp(-\alpha\ x)$

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
- Acquaintance chosen 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

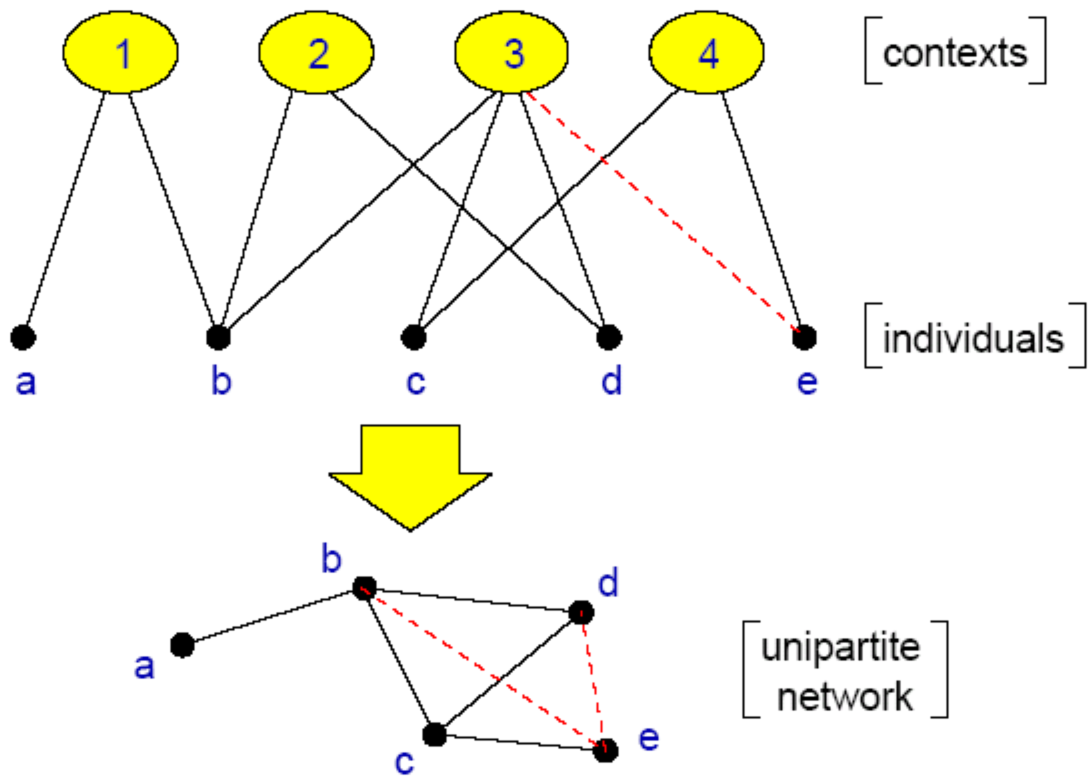
Successful chains disproportionately used
- weak ties (Granovetter)
- professional ties (34% vs. 13%)
- ties originating at work/college
- target's work (65% vs. 40%)

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

Strategy: Geography -> Work

Origins of small worlds: group affiliations

Social distance—Bipartite networks:



Origins of small worlds: other generative models

- Assign properties to nodes (e.g. spatial location, group membership)
- Add or rewire links according to some rule
  - optimize for a particular property (simulated annealing)
  - add links with probability depending on property of existing nodes, edges (preferential attachment, link copying)
  - simulate nodes as agents 'deciding' whether to rewire or add links

Origins of small worlds: efficient network example trade-off between wiring and connectivity

Small worlds: How and Why, Nisha Mathias and Venkatesh Gopal

$$E = \lambda L + (1-\lambda)W$$

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

$$W = \sum_{e_{ij}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

- E is the ‘energy’ cost we are trying to minimize
- L is the average shortest path in ‘hops’
- W is the total length of wire used

Origins of small worlds: efficient network example another model of trade-off between wiring and connectivity

physical distance hop penalty

$$\text{effective length of edge } (i,j) = \lambda\sqrt{n}\, d_{ij} + (1-\lambda)$$

- Incorporates a person's preference for short distances or a small number of hops
  - What do you think the differences in network topology will be for car travel vs. airplane travel?
- Construct network using simulated annealing

Air traffic networks

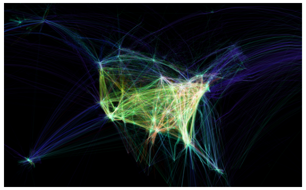
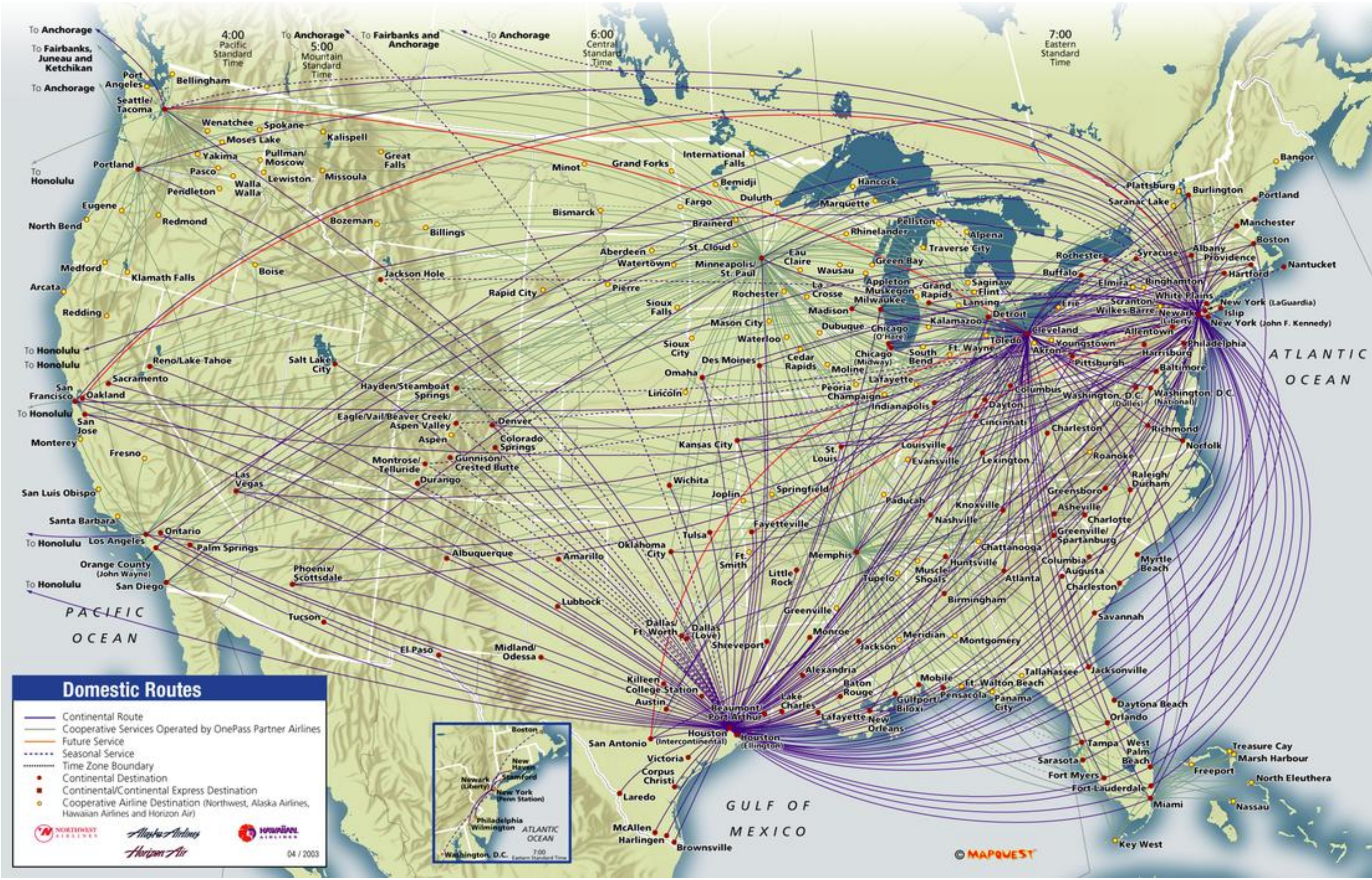
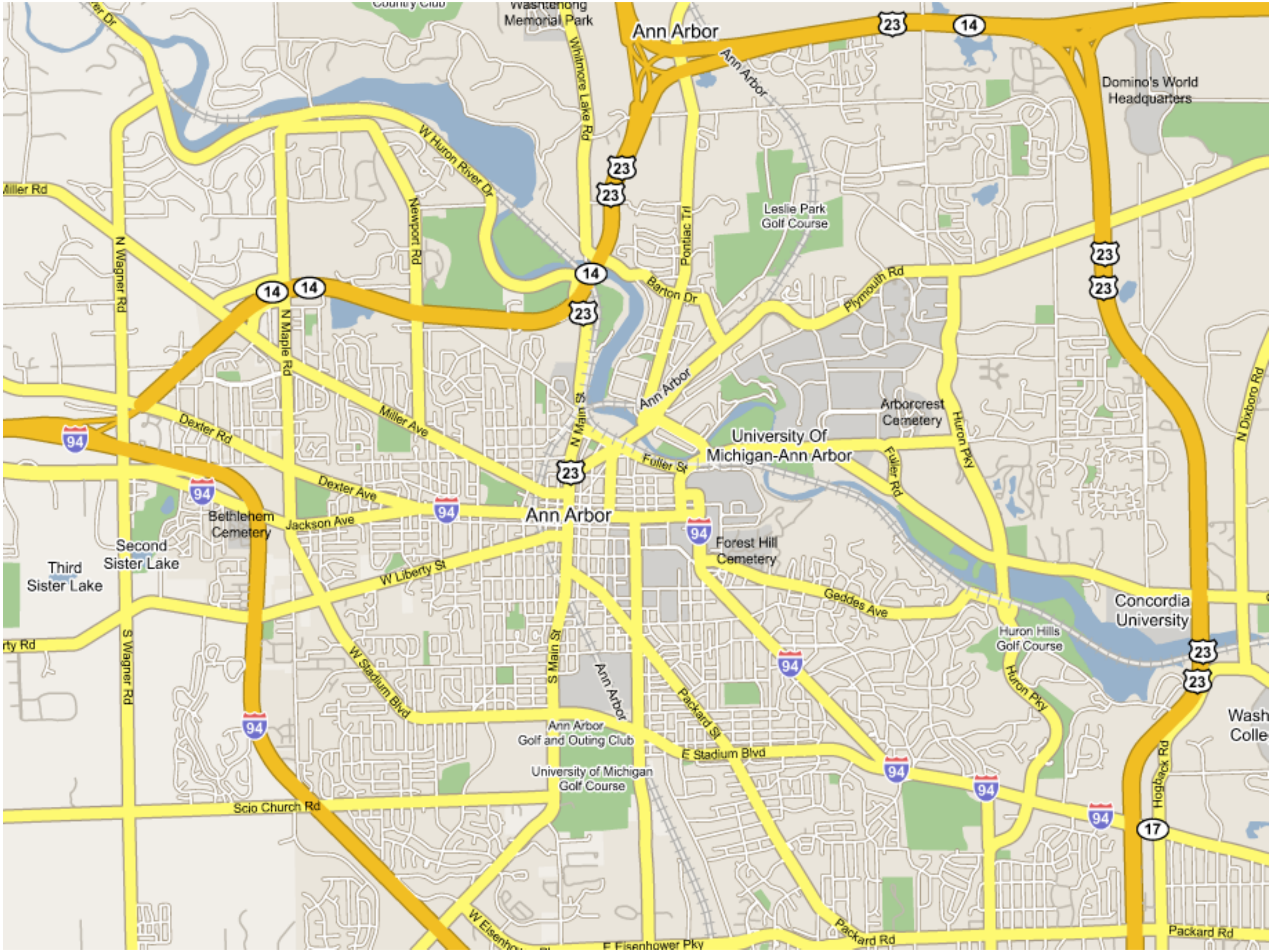


Image: Aaron Koblin

http://aaronkoblin.com/gallery/index.html



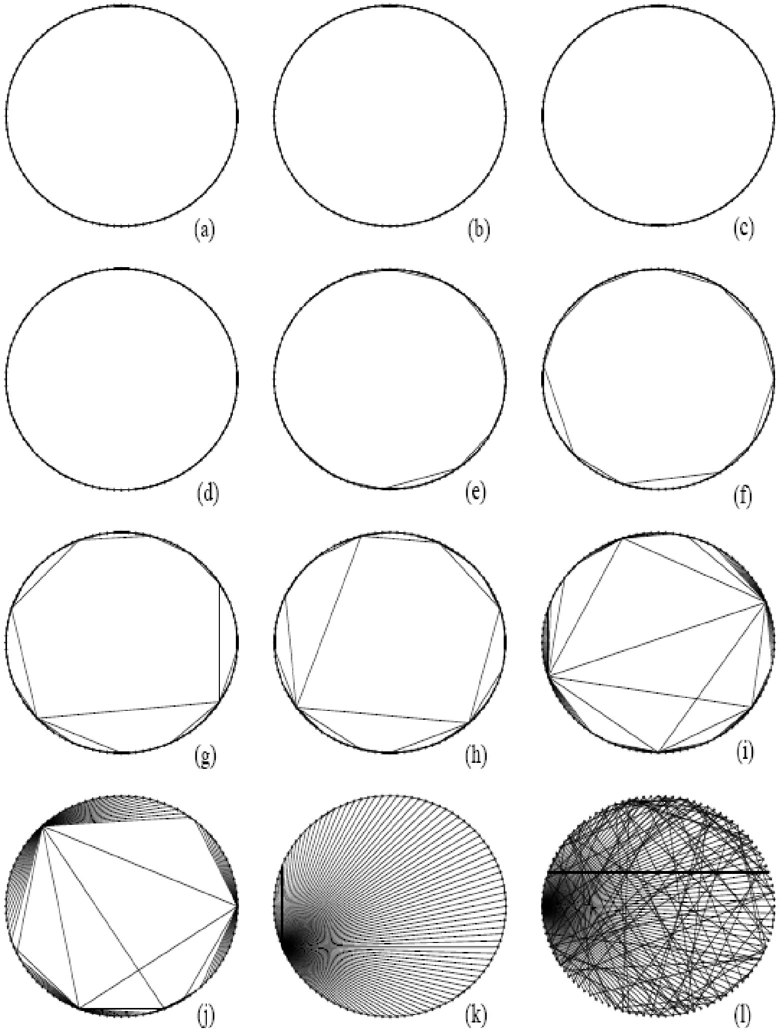
Source: Continental Airlines, http://www.continental.com/web/en-US/content/travel/routes/default.aspx



Source: http://maps.google.com

Origins of small worlds: tradeoffs

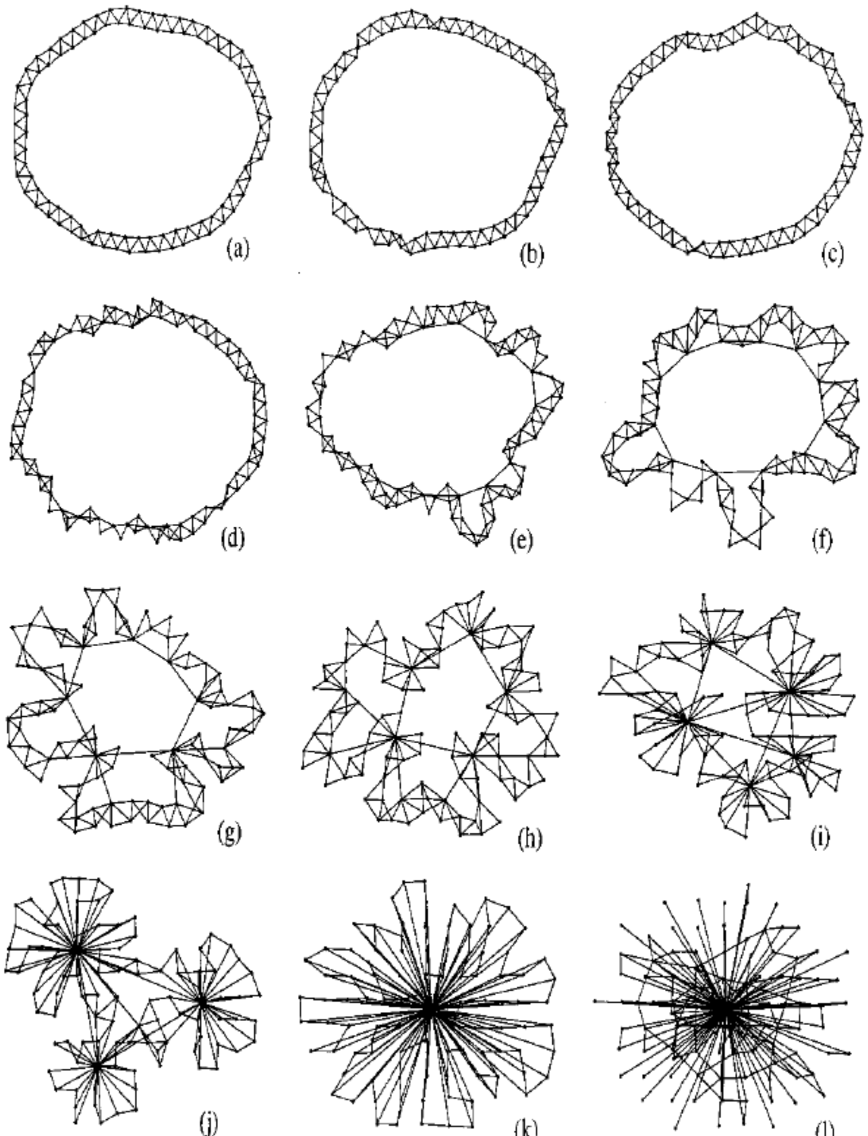
- rewire using simulated annealing
- sequence is shown in order of increasing λ



Source: Small worlds: How and Why, Nisha Mathias and Venkatesh Gopal

http://link.aps.org/doi/10.1103/PhysRevE.63.021117 DOI: 10.1103/PhysRevE.63.021117

Origins of small worlds: tradeoffs

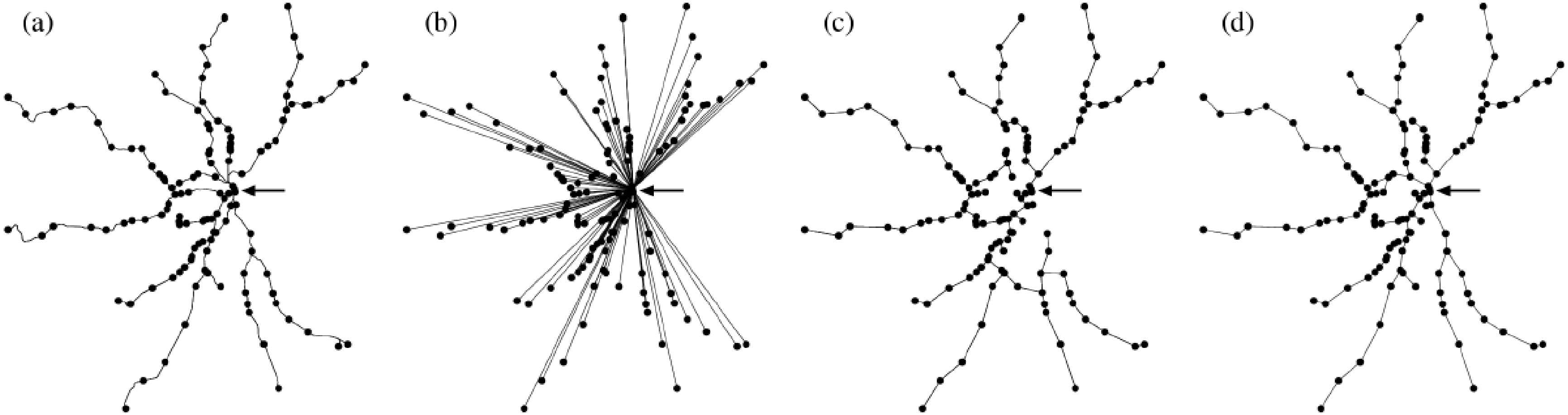


- same networks, but the vertices are allowed to move using a spring layout algorithm
- wiring cost associated with the physical distance between nodes

Source: Small worlds: How and Why, Nisha Mathias and Venkatesh Gopal

http://link.aps.org/doi/10.1103/PhysRevE.63.021117 DOI: 10.1103/PhysRevE.63.021117

Origins of small worlds: tradeoffs



(a) Commuter rail network in the Boston area. The arrow marks the assumed root of the network.

(b) Star graph.

(c) Minimum spanning tree.

(d) The model applied to the same set of stations.

add edge with smallest weight

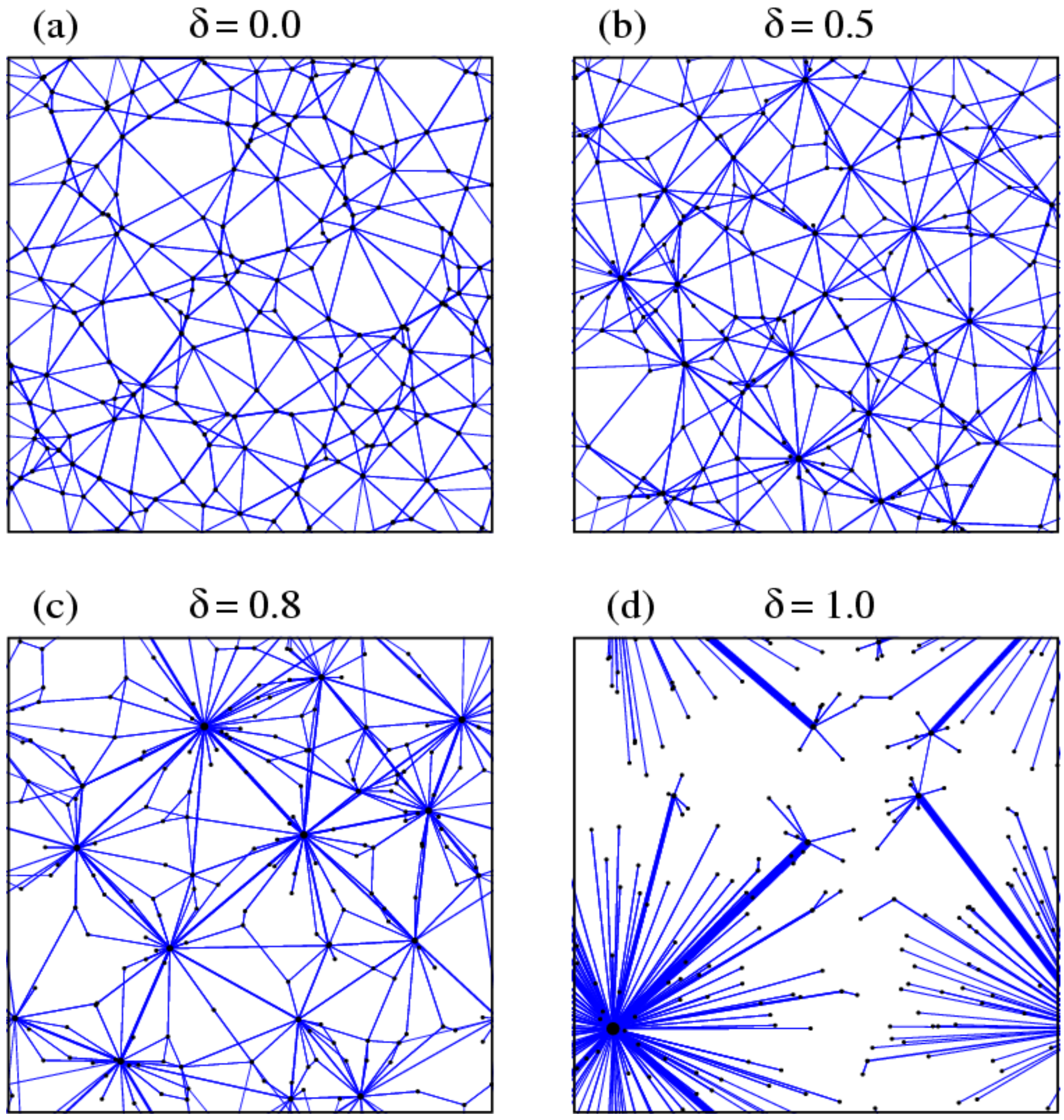
$$w'_{ij} = d_{ij} + \beta l_{j0}$$

hops to root node

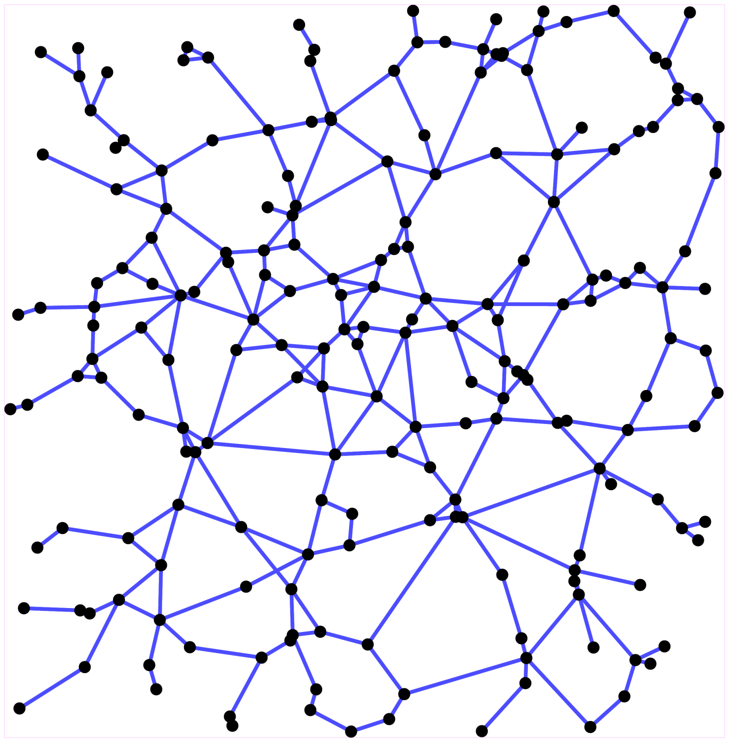
Euclidean distance between i and j

Source: Small worlds: How and Why, Nisha Mathias and Venkatesh Gopal

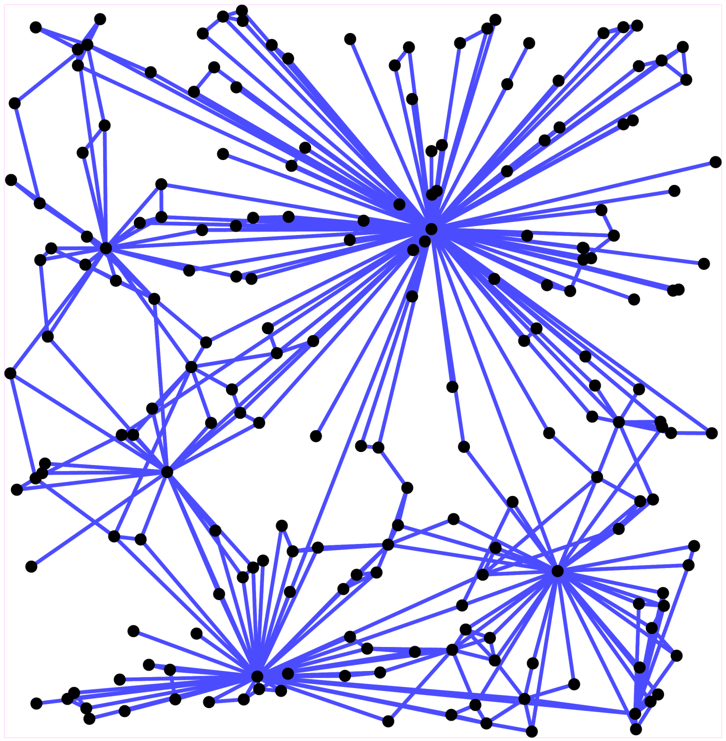
http://link.aps.org/doi/10.1103/PhysRevE.63.021117 DOI: 10.1103/PhysRevE.63.021117



Source: The Spatial Structure of Networks, M. T. Gastner and M. E.J. Newman
http://www.springerlink.com/content/p26t67882668514q DOI: 10.1140/epjb/e2006-00046-8



Roads



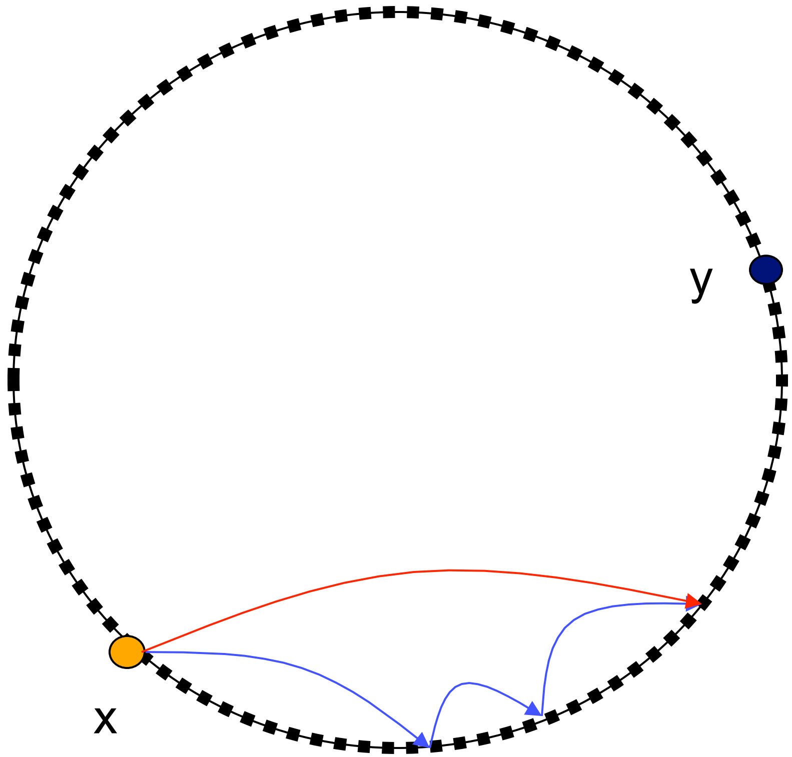
Air routes

Source: The Spatial Structure of Networks, M. T. Gastner and M. E.J. Newman
http://www.springerlink.com/content/p26t67882668514q DOI: 10.1140/epjb/e2006-00046-8

Origins of small worlds: navigation

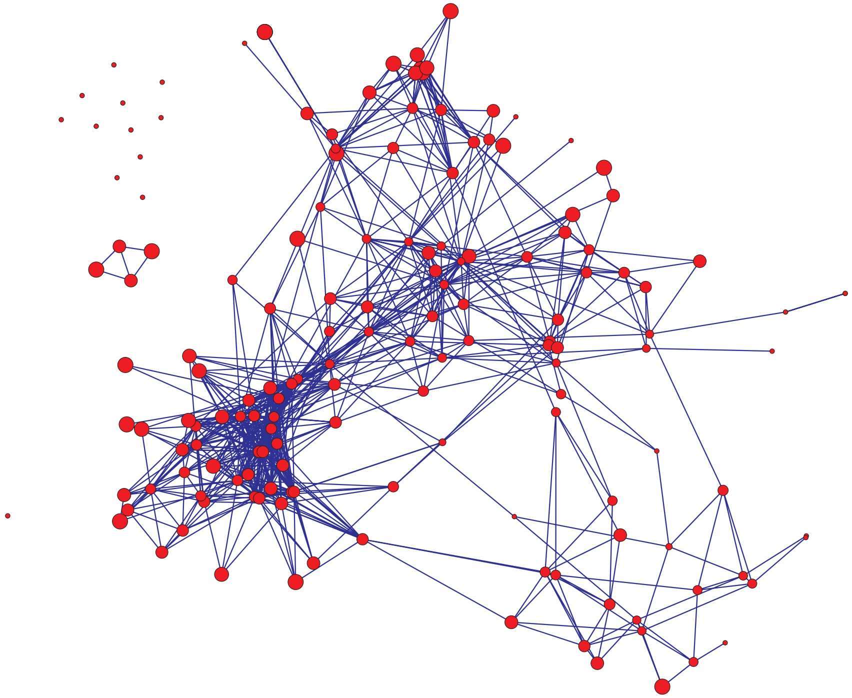
Aaron Clauset and Christopher Moore

arxiv.org/abs/cond-mat/0309415

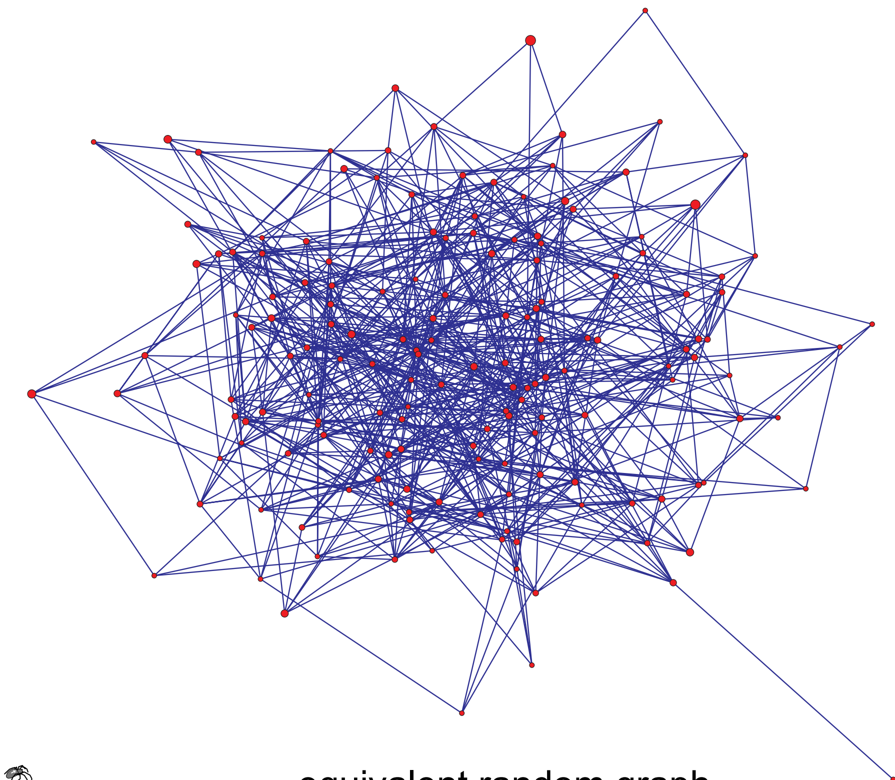


- start with a 1-D lattice (a ring)
- we start going from x to y, up to s steps away
- if we give up (target is too far), we rewire x's long range link to the last node we reached
- long range link distribution becomes 1/r, r = lattice distance between nodes
- search time starts scaling as log(N)

PS 3: is your network a small world?



Lada's Facebook network



equivalent random graph

nodes are sized by clustering coefficient

Small world networks: Summary

- The world is small!
- Watts & Strogatz came up with a simple model to explain why
- Other models incorporate geography and hierarchical social structure
- Small worlds may evolve from different constraints (navigation, constraint optimization, group affiliation)