Complex Network Theory

Lecture 6

Small world networks

Instructor: S. Mehdi Vahidipour
(Vahidipour@kashanu.ac.ir)

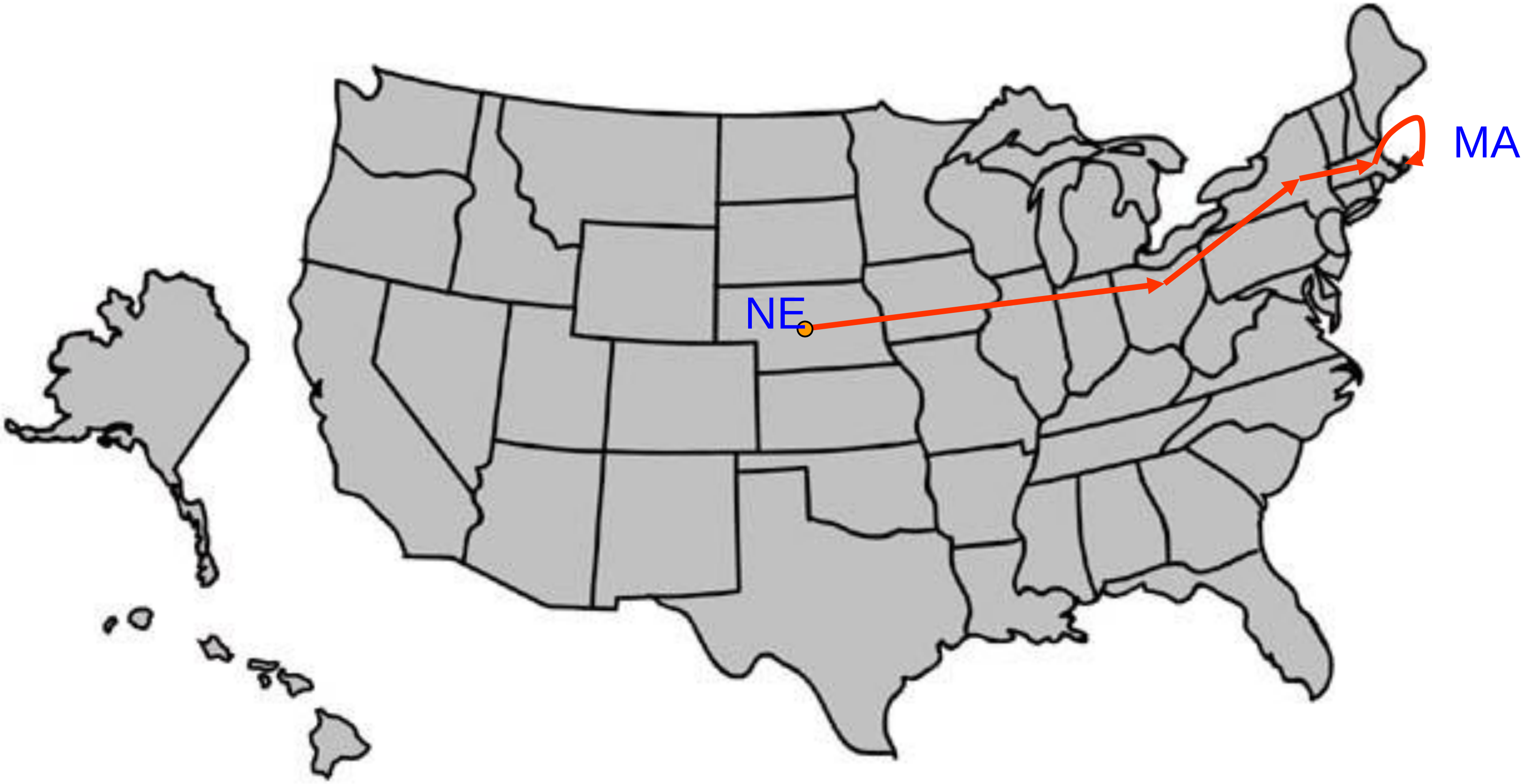
Thanks A. Rezvanian
A. Barabasi, L.Adamic,

Outline

- Milgram's small world experiment
- Watts & Strogatz small world model
- Kleinberg small world model
- Watts, Dodds & Newman community model
- Network models: a few examples

- Next class
  - Scale free networks

Small world experiments then

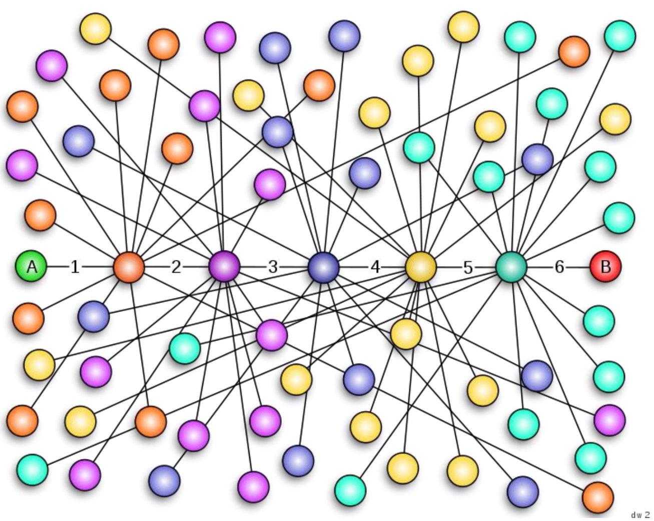


Milgram's experiment (1960's):

Given a target individual and a particular property, pass the message to a person you correspond with who is "closest" to the target.

Small-world experiment

- Start
  - Omaha, Nebraska, and Wichita, Kansas
- End
  - Boston, Massachusetts



One possible path of a message in the "Small World" experiment by Stanley Milgram.

Milgram's experiment

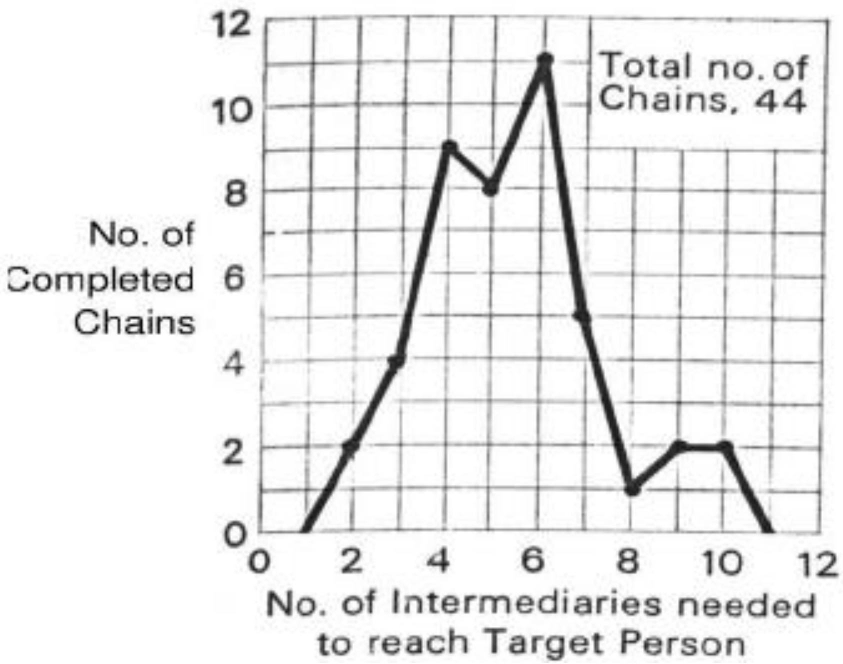
- Instructions:
  - Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target.
  - some letters: From Wichita (Kansas) and Omaha (Nebraska) to Sharon (Mass)
  - If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder to a personal acquaintance who is more likely than you to know the target person.
- Outcome:
  - 20% of initiated chains reached
  - Target average chain length = 6.5
  - "Six degrees of separation

Milgram, Psych Today **2**, 60 (1967)

Milgram's small world experiment

- Target person worked in Boston as a stockbroker.
- 296 senders from Boston and Omaha.
- 232 of the 296 letters never reached the destination
- 64 letters (20%) of senders reached target.
- average path length = 6.5.
- "Six degrees of separation"
- The **Small World** concept

in simple terms describes the fact despite their often **large size**, in most networks there is a **relatively short path between any two nodes.**



Degrees of separation in real networks

- In 2001, Watts attempted to recreate Milgram's experiment on the internet, using an e-mail message as the "package" that needed to be delivered, with 48,000 senders and 19 targets (in 157 countries). Watts found that the average (though not maximum) number of intermediaries was around 6.
- A 2007 study by Leskovec and Horvitz examined a data set of instant messages composed of 30 billion conversations among 240 million people. They found the average path length among Microsoft Messenger users to be 6.6 (some now call the theory, "the seven degrees of separation" because of this)

Small world experiments now

Email experiment by Dodds, Muhamad, Watts, Science 301, (2003)

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

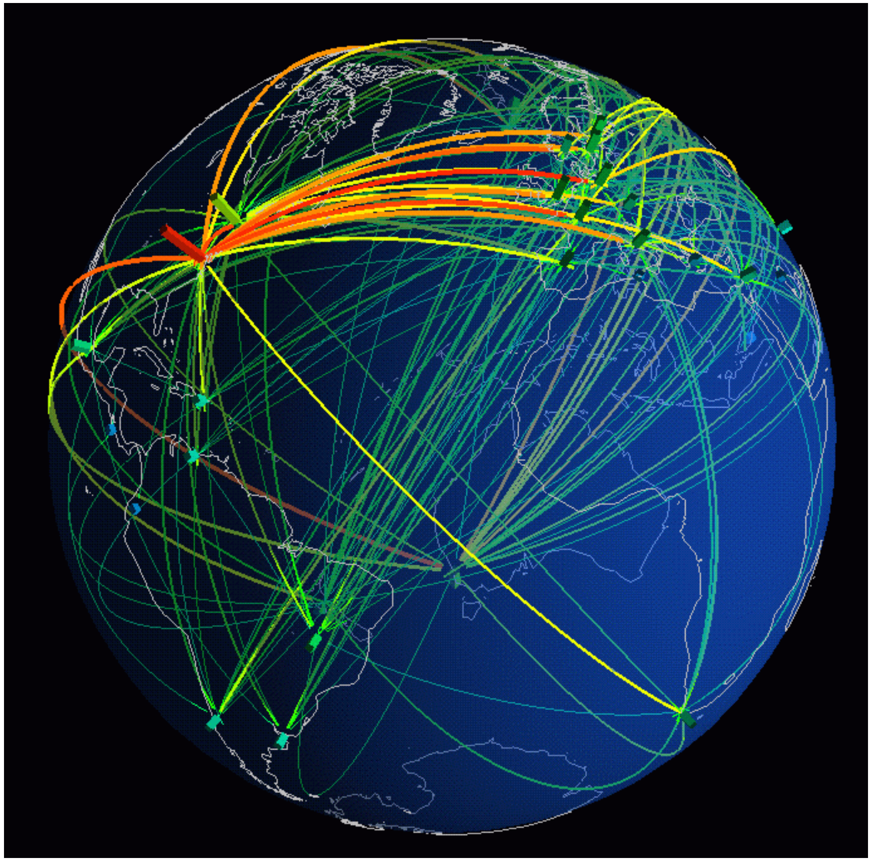


image by Stephen G. Eick
http://www.bell-labs.com/user/eick/index.html
(unrelated to small world experiment…)

Degrees of separation in real networks

- Species in food webs appear to be on average 2 links away from each other.
- Molecules in the cell are separated on average by 3 chemical reactions.
- Scientists in different fields of science are separated by 4 to 6 co-authorship links.
- The neurons in the brain of the C. elegans worm are separated by 14 synapses.
- The Web holds the absolute highest record of 20 to 22
- The Internet, a network of hundreds of thousands of routers, has a separation of 10 to 12

Interpreting Milgram's experiment

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

High node degrees in real networks

- How do networks achieve such a uniformly short path despite consisting of billions of nodes?
- The answer lies in the highly interconnected nature of these networks.
- Why in real networks, nodes have many more links than one (the threshold for connectivity)?
- At the critical point when the average connectivity is around one per node, the separation between nodes could be rather large.
- But as we add more links, the distance between the nodes suddenly collapses.

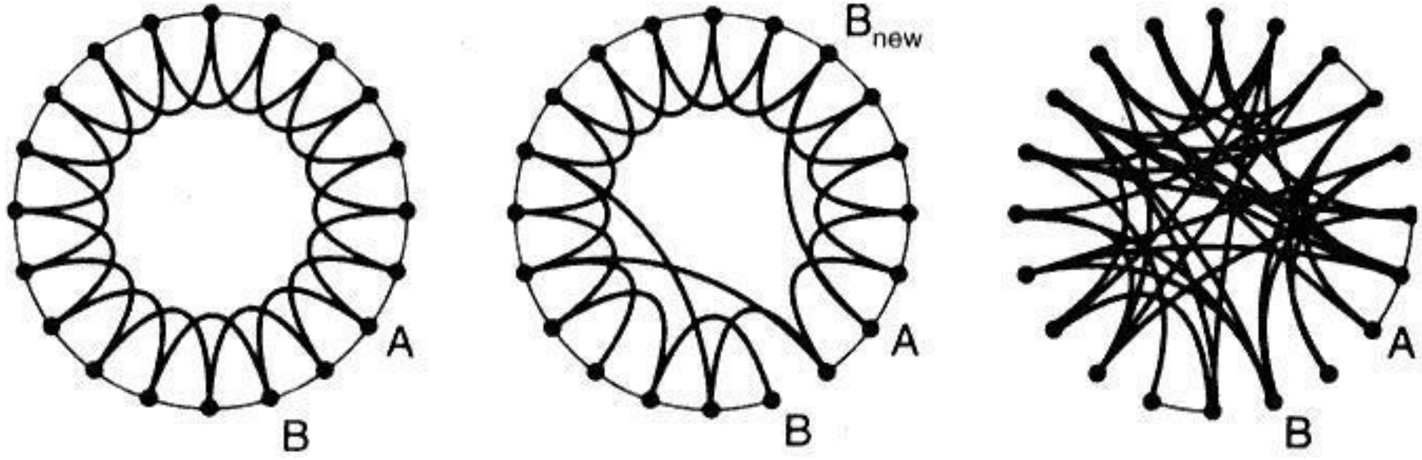
Small-world networks

- Watts and Strogtaz showed that many real-world networks:
  - Have small characteristic path length compared to random networks
  - At the same time, have high clustering coefficient that is much larger than that of random networks
  - There are indeed small-worlds
- This discovery had huge impact on the various developments in Network fields
  - Search in complex networks
  - Communication in networks
  - Synchronization and consensus

The small world model

High clustering: my friends' friends tend to be my friends

Watts & Strogatz (1998) - a few random links in an otherwise clustered graph give an average shortest path close to that of a random graph



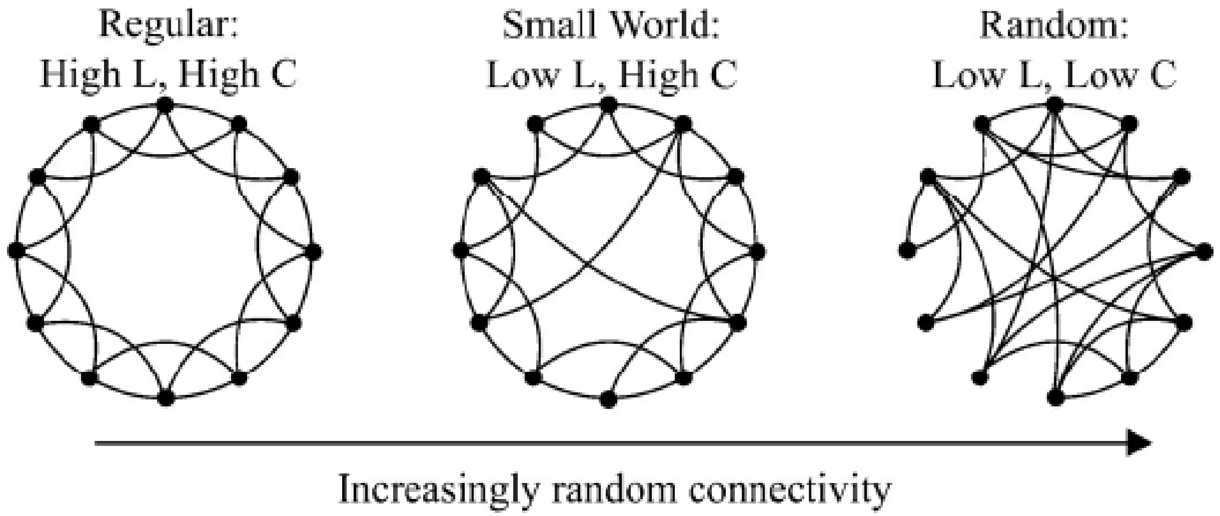
Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

Watts-Strogatz model

- The construction algorithm:
  - Consider a ring graph where each node is connected to its **k** nearest neighbors with undirected edges (k-regular)
  - Choose a node and one of the edges that connects it to its nearest neighbors and then with probability **P** reconnect this edge to a node randomly chosen over the graph
    - provided that the duplication of edges and self-loops are forbidden
  - The process is repeated until all nodes and nearest neighbor connecting edges are met
  - Next, the edges that connect the nodes to their second nearest neighbors are reconnected and the rewiring process is performed on them with the same conditions as above
  - The same procedure is then repeated for the remaining edges connecting the nodes to their **k** nearest neighbors

Watts-Strogatz model (WS model)

- The resulting graph is so that
  - for the value of **P = 0** we will have the original ring graph
  - for the value of **P = 1** produces a pure **random graph**
  - For some values of **P** between these two extremes the resulting network has small characteristics path length, and at the same time, high clustering coefficient
- The average degree will be <k> = k , edges: nk/2



Networks in nature (empirical observations)

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

$$l_{\text{network}} \approx \ln(N)$$

$$C_{\text{network}} >> C_{\text{randomgraph}}$$

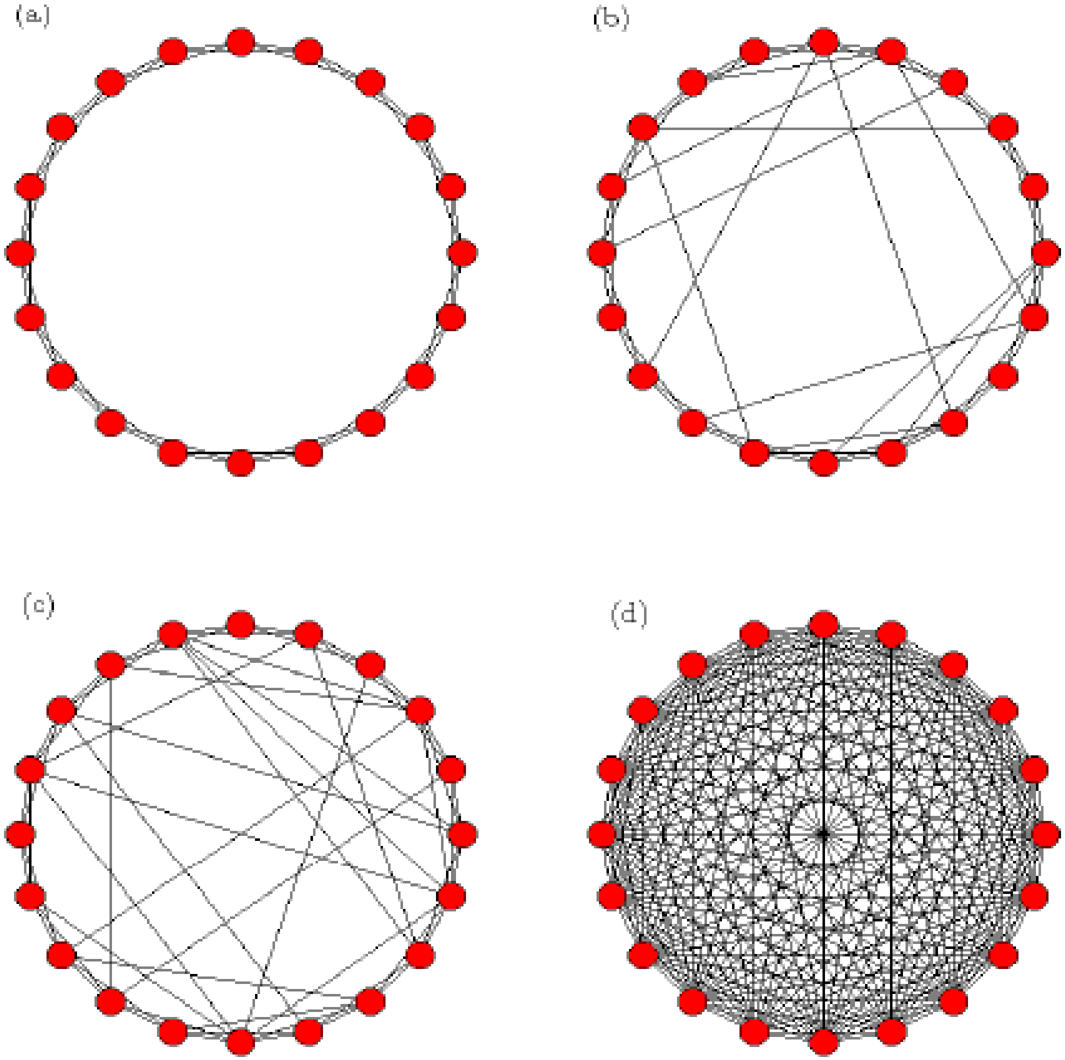
Newman-Watts model

- Starting with a k-ring graph
- N nodes
- Non-connected nodes get connected with probability P
- P = 1 results in complete graph
- for some small values of P
  - small-world property
  - high transitivity
- The networks are always connected

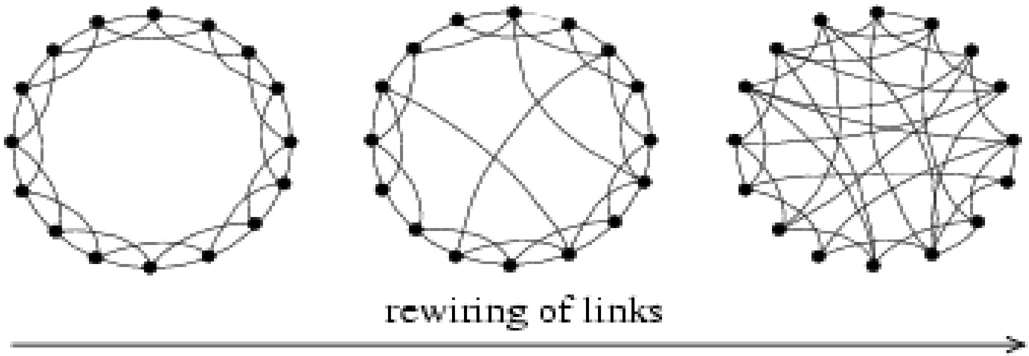
Newman-Watts model

20 nodes in a 2-regular ring with

- a) P = 0
- b) P = 0.05
- c) P = 0.15
- d) P = 1

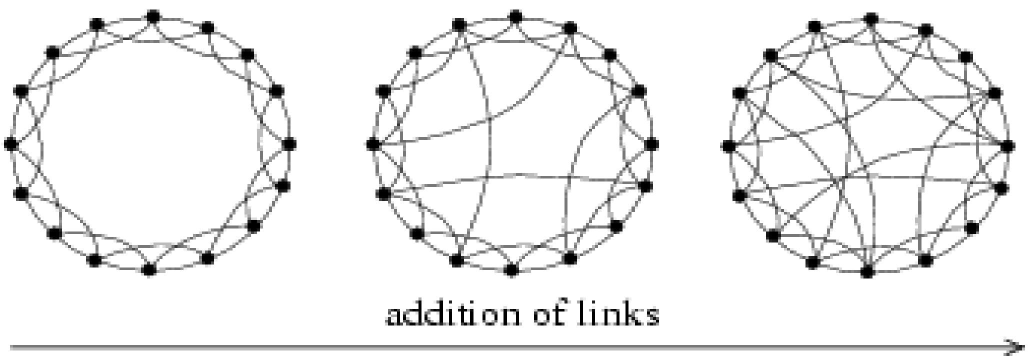


Two ways of constructing a small world graph



Select a fraction p of edges
Reposition on of their endpoints

(Watts-Strogatz model)



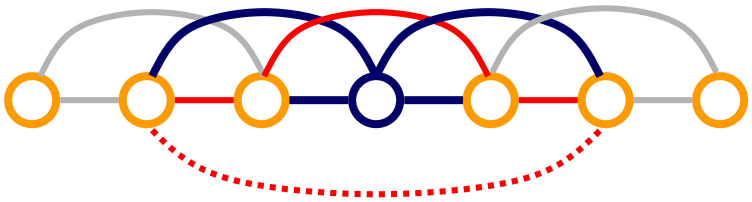
Add a fraction p of additional edges leaving underlying lattice intact

(Newman-Watts model)

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

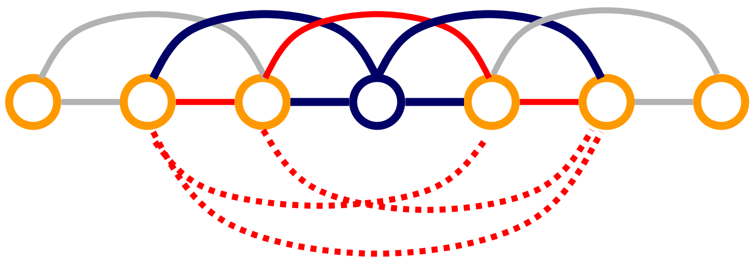
Original model

- Each node has K nearest neighbors (local)
- Probability **p** of rewiring to randomly chosen nodes
- p small: regular lattice
- p large: classical random graph



p=0 Ordered lattice

- Compute the clustering coefficient as follows
  - each node is connected to K neighbors, who can have K*(K-1)/2 pairwise connections between them
  - some of the connections between them are present in the lattice



Caution: sometimes the lattice will be specified as

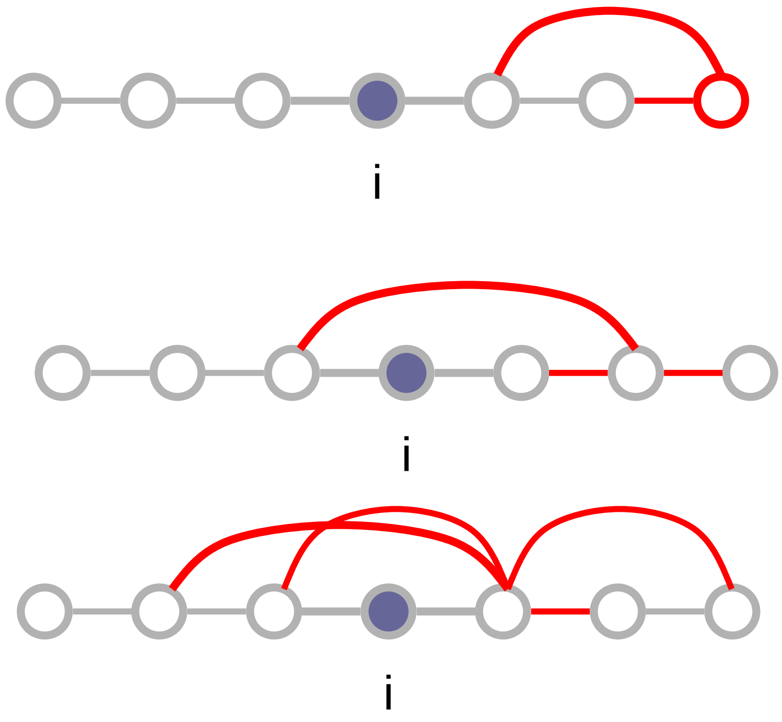
each node connects to K closest neighbors

each node connects to all neighbors within distance k (k = K/2)

Clustering coefficient for regular lattice

- In general, can have any K
- a neighbor K/2 hops away from i can connect to (K/2 – 1) of i's neighbors
- a neighbor K/2-1 hops away can connect to (1 + K/2 – 1) neighbors
- K/2 – 2 hops away
  - (2 + K/2 – 1) neighbors
- 1 hop away
  - 2*(K/2 – 1)

- Sum this up
  - multiply by factor of 2 because i has neighbors on both sides
  - divide by a factor of 2 because edges are undirected



Clustering coefficient for regular lattice

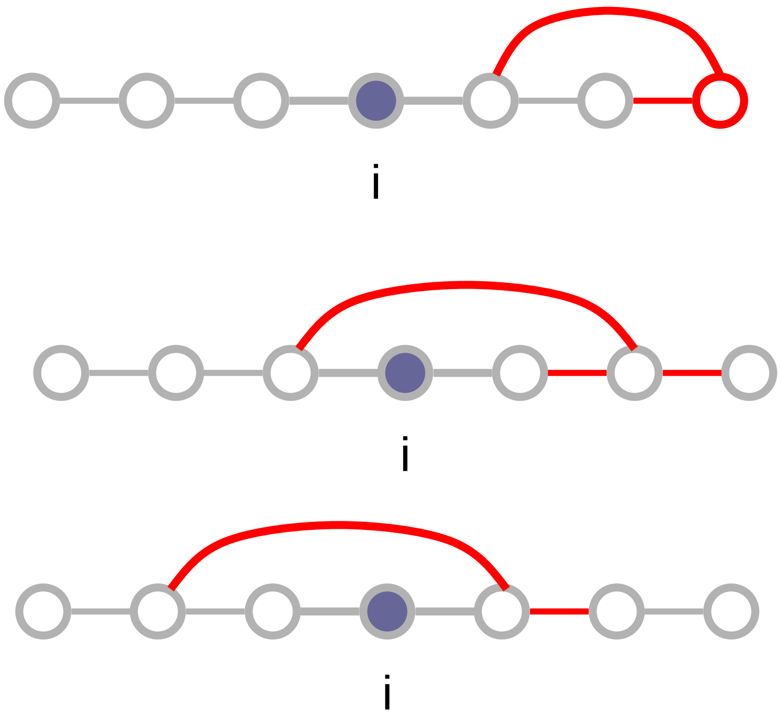
- The number of connections between neighbors is given by

$$\sum_{j=0}^{\frac{K}{2}-1}\left(\frac{K}{2}+i-1\right)=\frac{3}{8}K(K-2)$$

- The maximum number of connections is K*(K-1)/2

- → clustering coefficient is

$$C=\frac{3(K-2)}{4(K-1)}$$



Average shortest path – regular lattice

- Average node is N/4 hops away (a quarter of the way around the ring), and you can hop over K/2 nodes at a time

$$l \approx \frac{N}{2K} >> 1$$

p=1 Random graph

$$l \approx \frac{\ln N}{\ln K} \qquad \text{small}$$

$$C \approx \frac{K}{N} \qquad \text{small}$$

There are an average of K links per node.

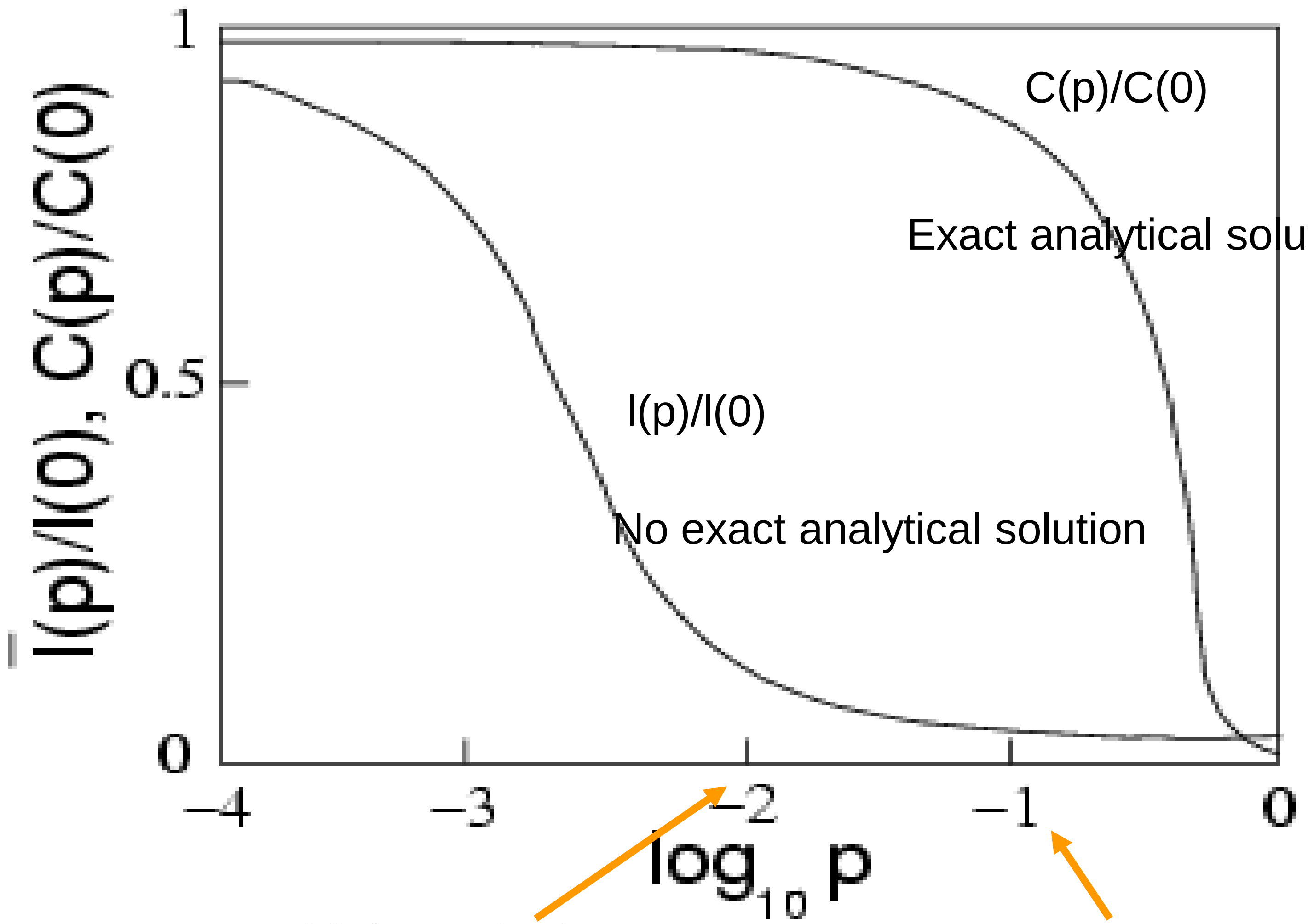
The probability that any two nodes are connected is p = K/N.

The probability that two nodes which share in a neighbor in common are connected themselves is the same as any two random nodes: K/N (actually (K-1)/N because they have already expended one edge on their common neighbor.

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

Change in clustering coefficient and average path length as a function of the proportion of rewired edges



N=1000

K=10

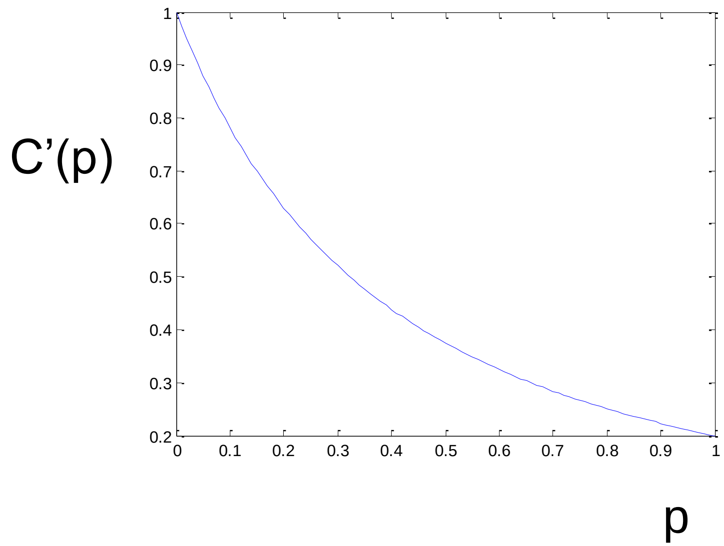
1% of links rewired

10% of links rewired

Clustering coefficient: addition of random edges

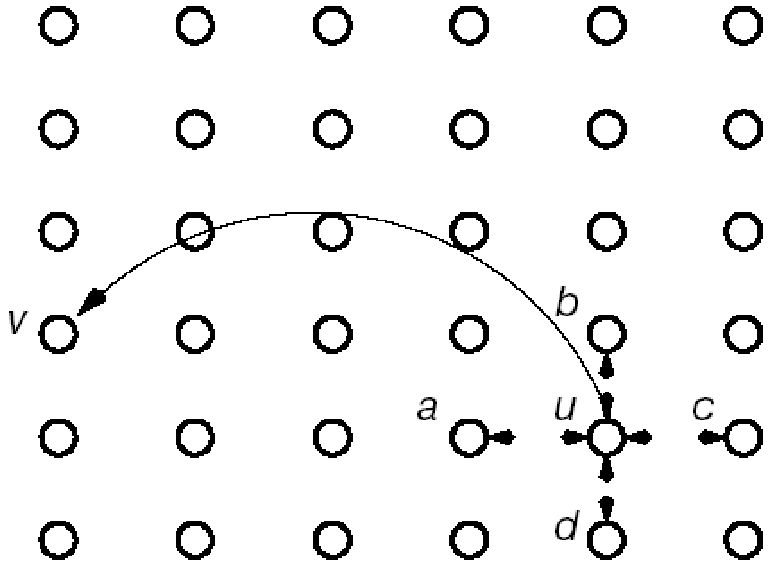
- How does C depend on p?
- C'(p)= 3 × number of triangles / number of connected triples
- C'(p) computed analytically for the small world model without rewiring

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



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

Kleinberg's geographical small world model



nodes are placed on a lattice and connect to nearest neighbors

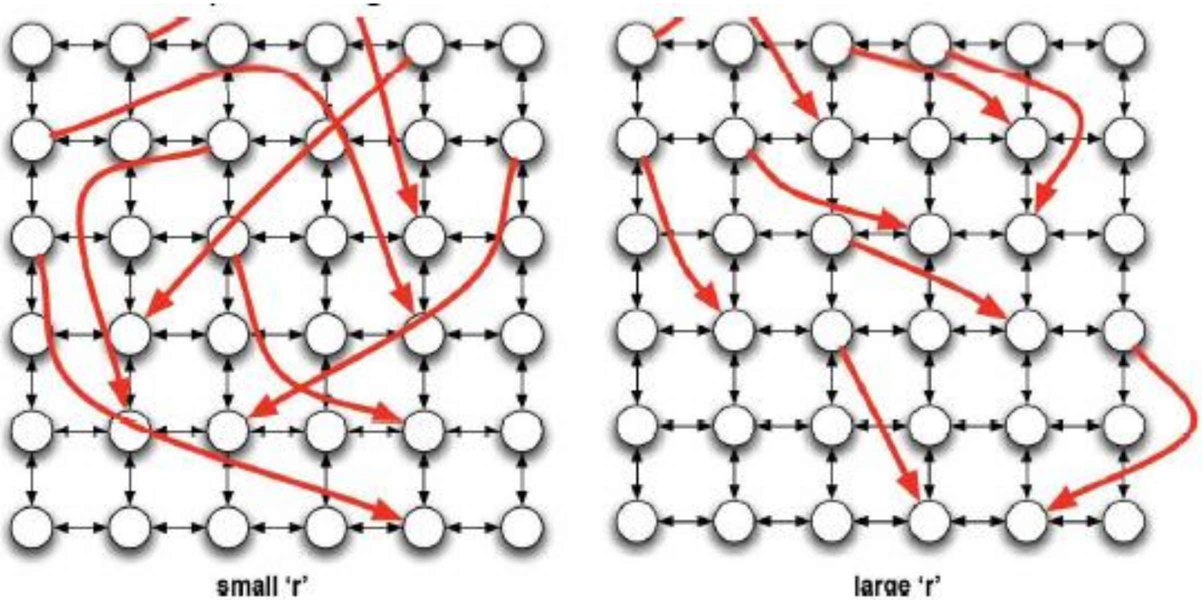
exponent that will determine navigability

additional links placed with $p_{uv} \sim \; d_{uv}^{-r}$

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

Small-worlds: algorithmic view

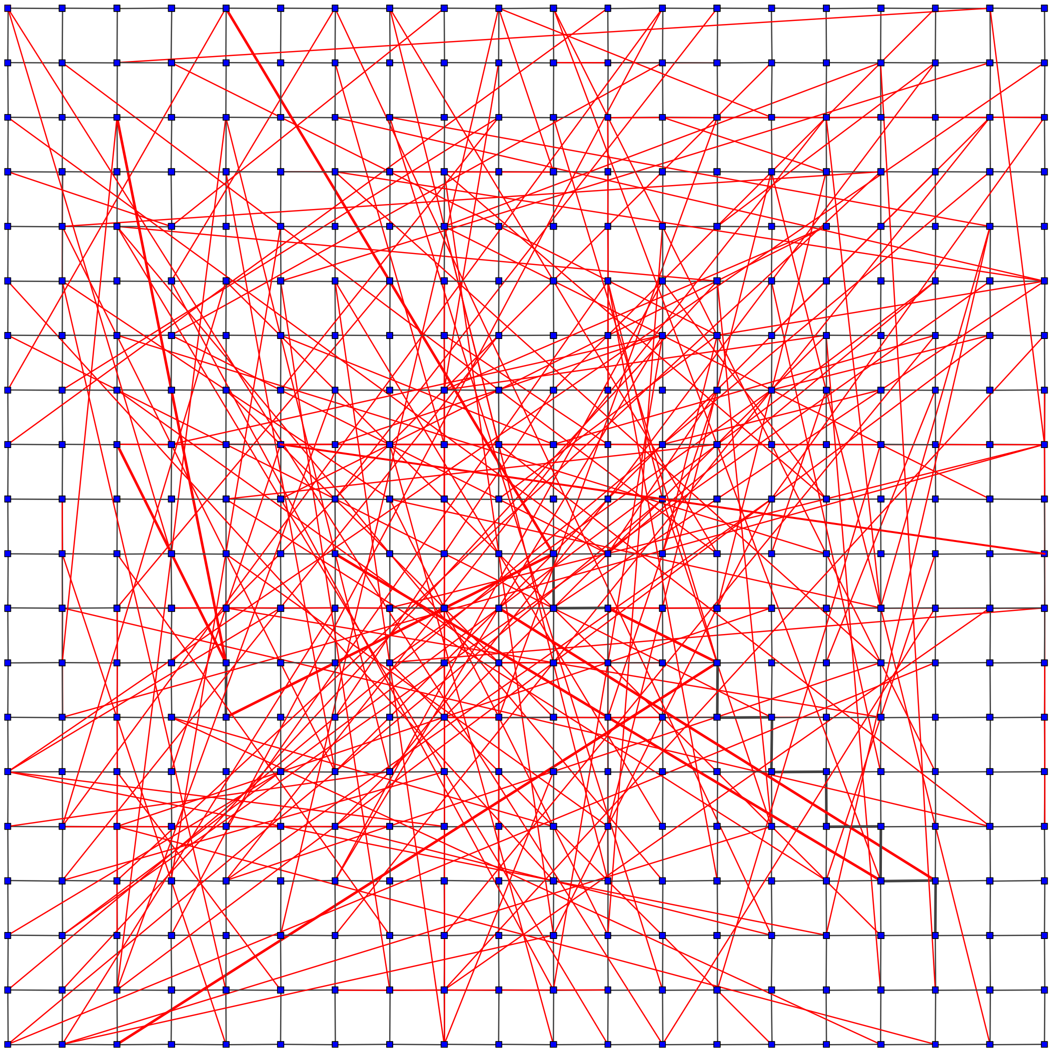
- Probability [u has v as its long range contact] :
- Infinite family of networks:
  - r = 0: each node's long-range contacts are chosen independently of its position on the grid
  - As r increases, the long range contacts of a node become clustered in its vicinity on the grid



no locality

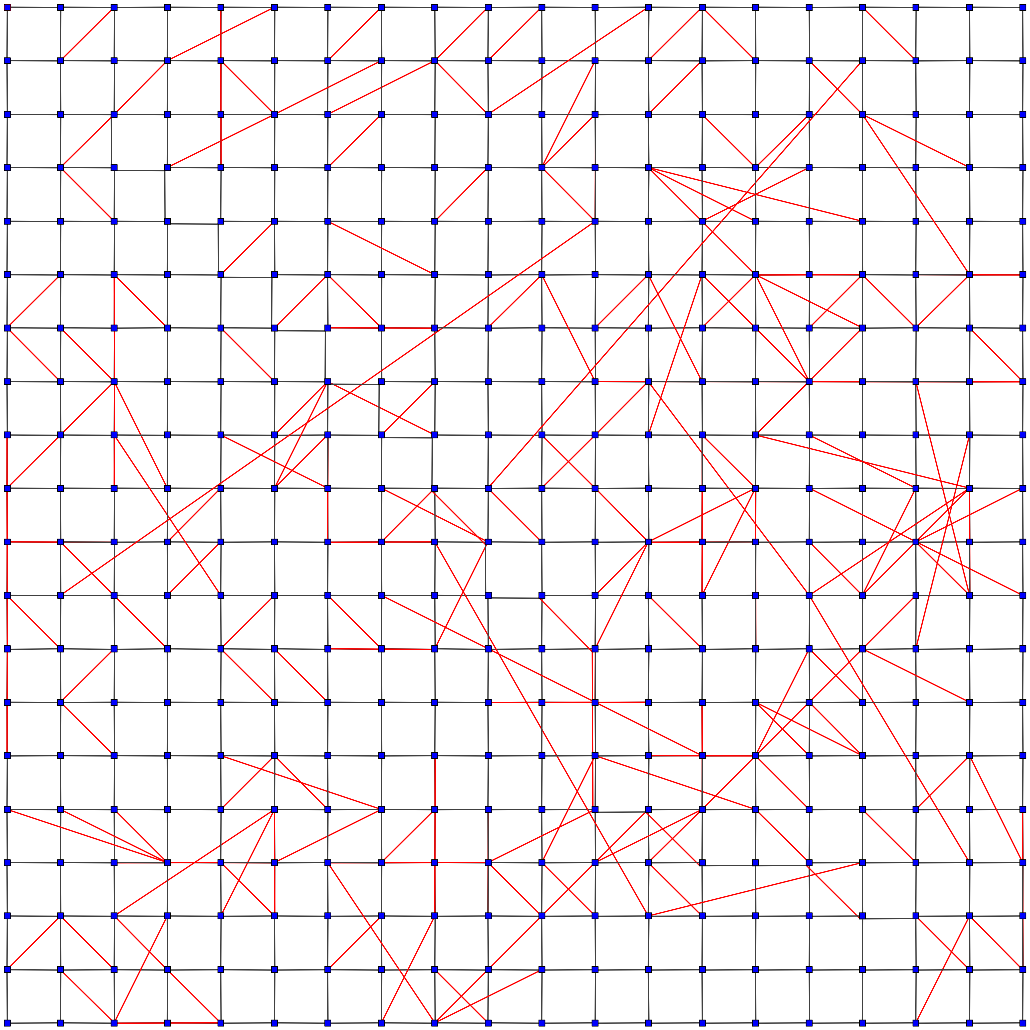
When **r=0**, links are randomly distributed, ASP ~ **log(n)**, n size of grid

$p \sim p_0$



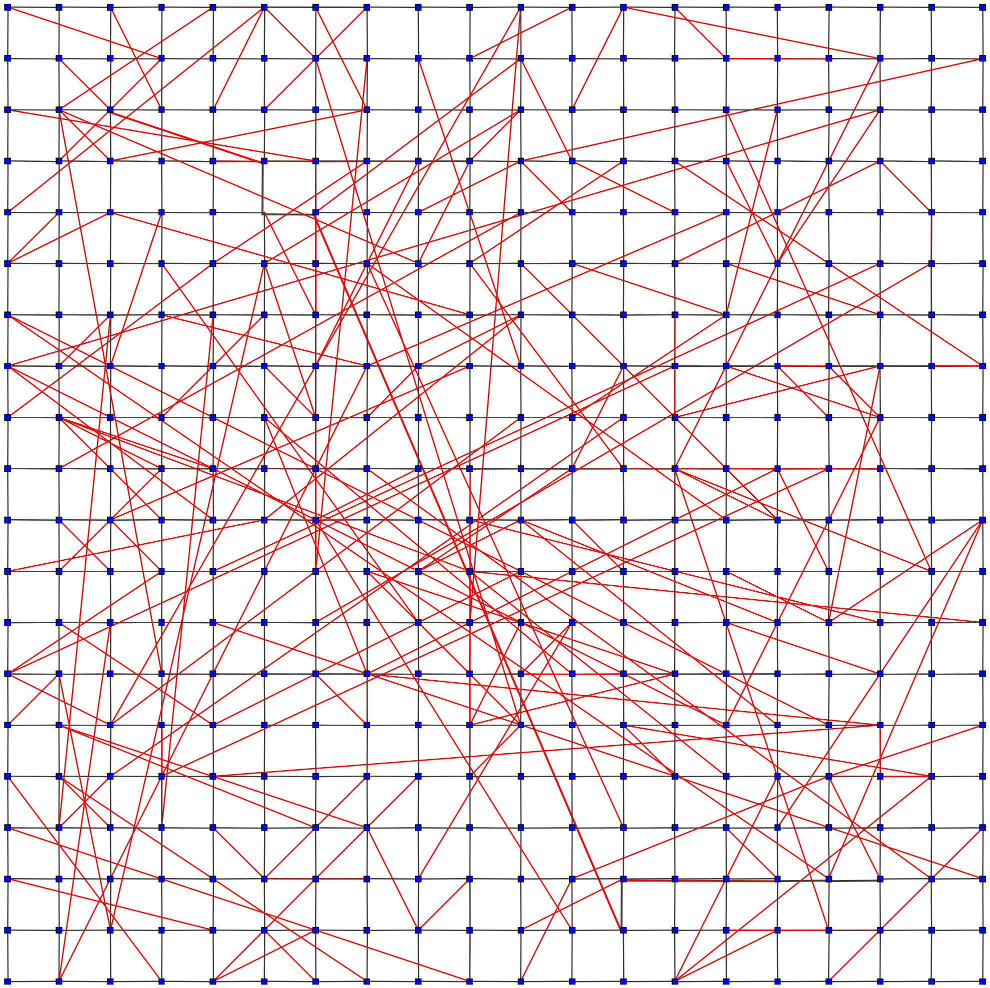
Links highly localized links on a lattice

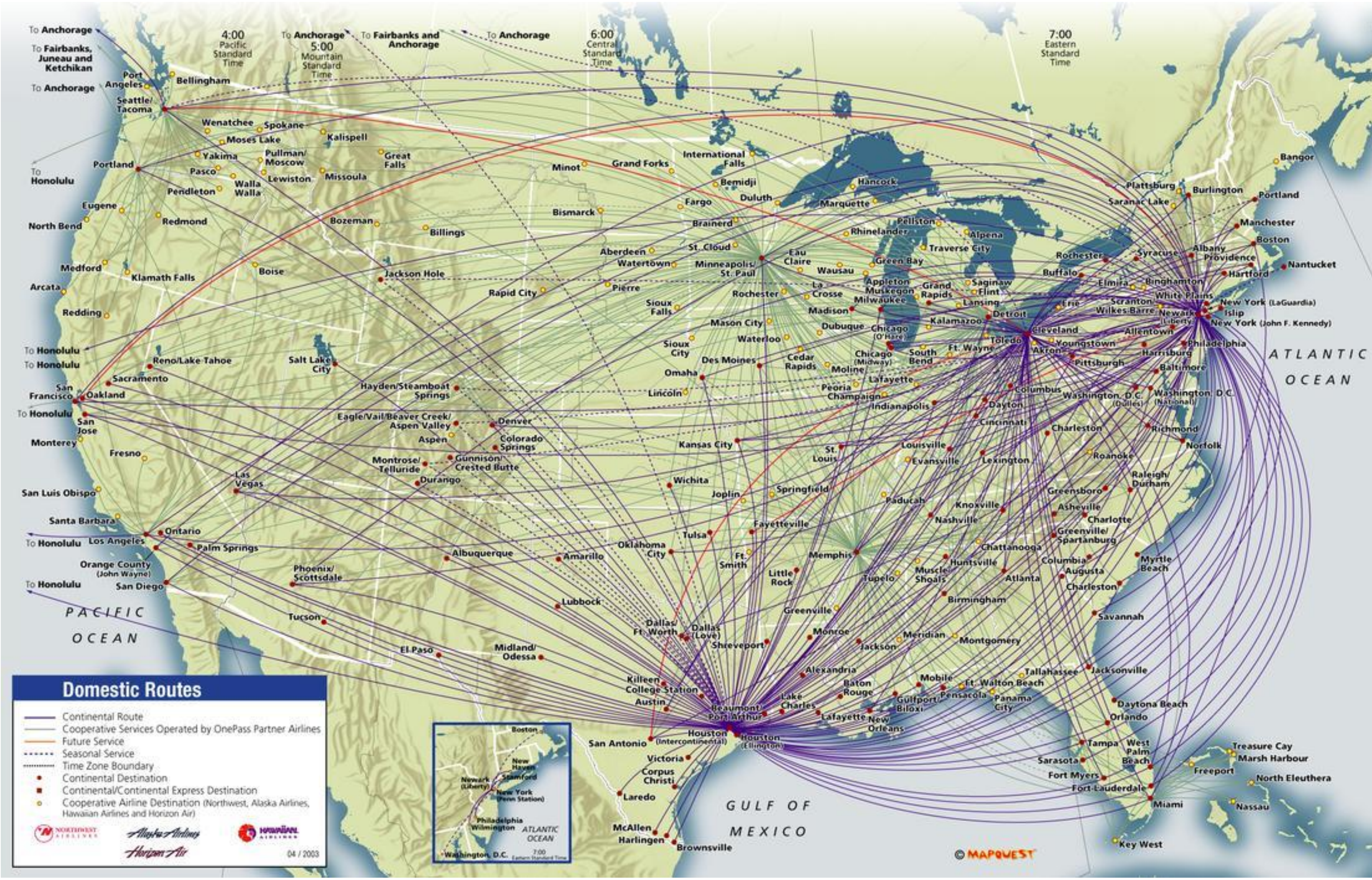
$$p \sim \frac{1}{d^4}$$



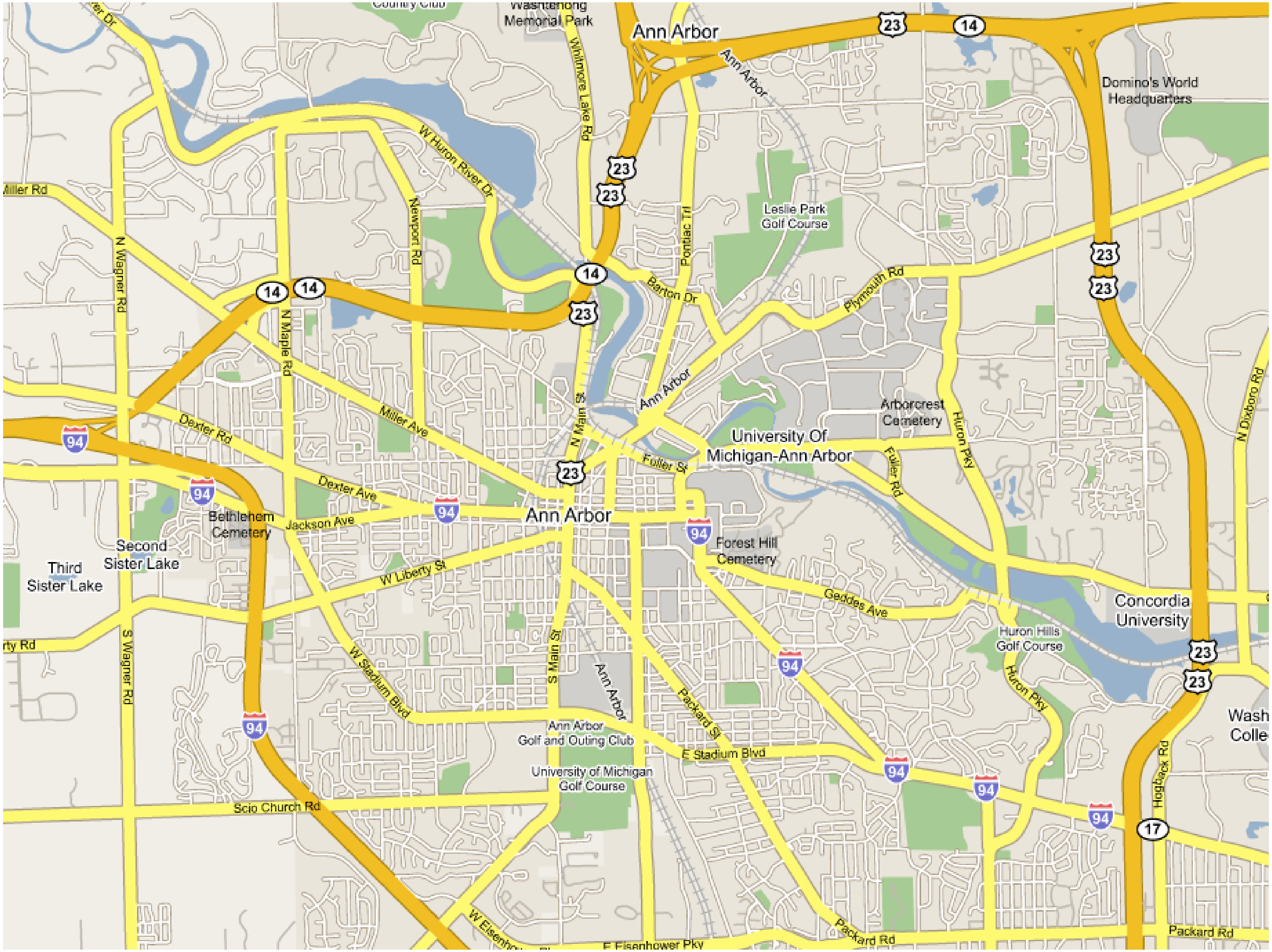
Links balanced between long and short range

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

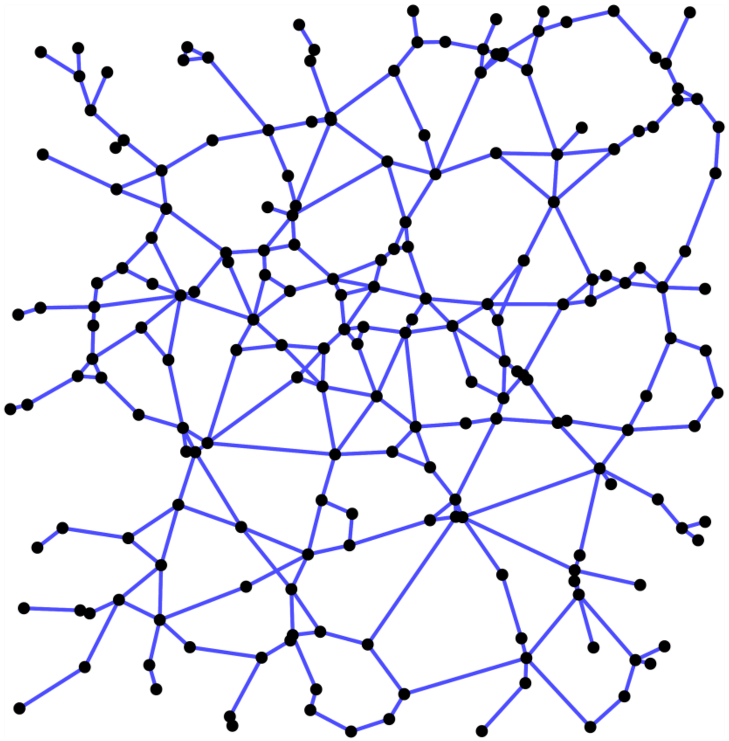




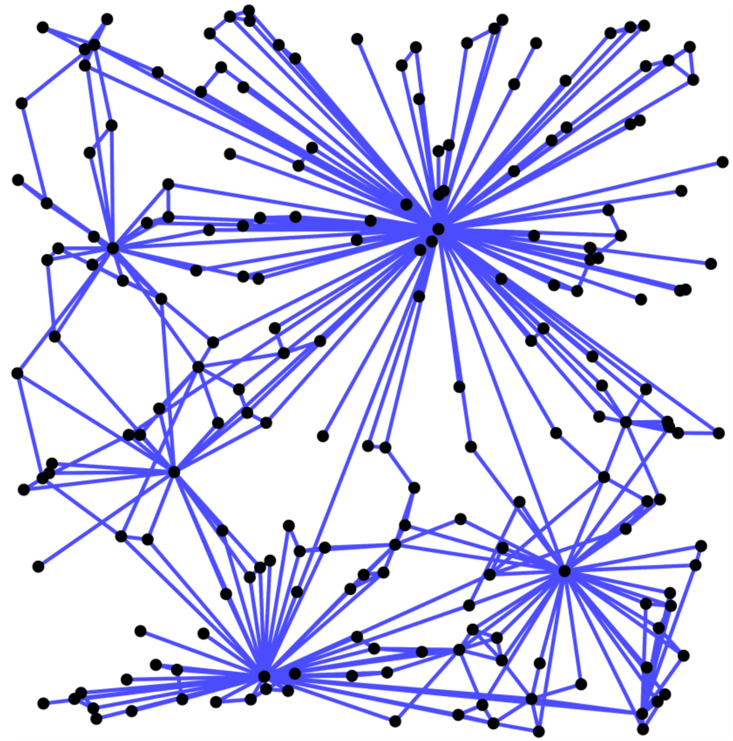
slide by Mark Newman



slide by Mark Newman



Roads



Air routes

slide by Mark Newman

Summary

- The world is small!
- Watts & Strogatz came up with a simple model to explain why
- Later, more sophisticated models of social structure were developed
- There are many, many more models that can be thought up and that give useful insights

Readings

- Newman, Mark. **Networks: an introduction**. Oxford University Press, 2010. (Chapter 15)
- Van Steen, Maarten. "**Graph Theory and Complex Networks** An Introduction, 2010. (Chapter 7)
- Easley and Kleinberg "**Networks, Crowds, and Markets**" (Chapters 20)
- Newman, Mark EJ. "**Random graphs as models of networks**." *Handbook of Graphs and Networks: From the Genome to the Internet* (2006).
- Watts DJ, Strogatz SH (1998) **Collective dynamics of 'small-world' networks**. Nature 393:440-442.
- Newman MEJ, Watts DJ (1999) **Renormalization group analysis of the small-world network model**. Physics Letters A 263: 341-346