Small World Networks

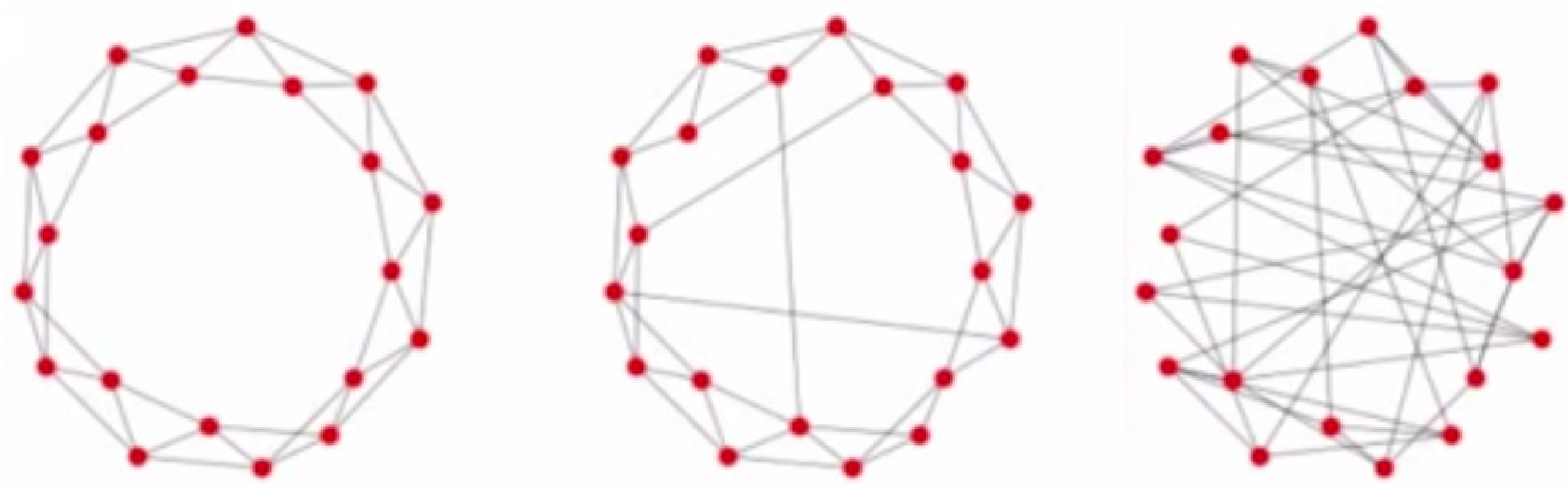
Dr. Natarajan Meghanathan
Associate Professor of Computer Science
Jackson State University, Jackson, MS
E-mail: natarajan.meghanathan@jsums.edu

Small-World Networks

- A small-world network is a type of graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops.
- Specifically, a small-world network is defined to be a network where the typical distance L (the number of hops) between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes in the network.
- Examples of Small-World Networks:
  - Road maps, food chains, electric power grids, metabolite processing networks, networks of brain neurons, voter networks, telephone call graphs, gene regulatory networks.

Small Worlds

- Two major properties of small world networks
  - High average clustering coefficient
    - The neighbors of a node are connected to each other
    - Nodes' contacts in a social network tend to know each other.
  - Short average shortest path length
    - Shorter paths between any two nodes in the network



(Regular graph) (Small-world network) (Random graph)

0 **Randomness** 1

Small Worlds

- Note that for the same number of nodes and edges, we could create:
  - Random graphs (with edges arbitrarily inserted between any two nodes) and
  - Regular graphs (with some specific pattern of how edges are inserted between nodes)
- Regular graphs tend to have relatively high average clustering coefficient
- Random graphs tend to have relatively low average shortest path length
- We could bring the best of the two graphs by generating a small world network as follows:
  - Remove a small fraction of the edges in a regular graph and re-insert them between any two randomly chosen nodes. This will not appreciably affect the average clustering coefficient of the nodes; but would significantly lower the average lengths of the shortest paths.

Small-World Experiment

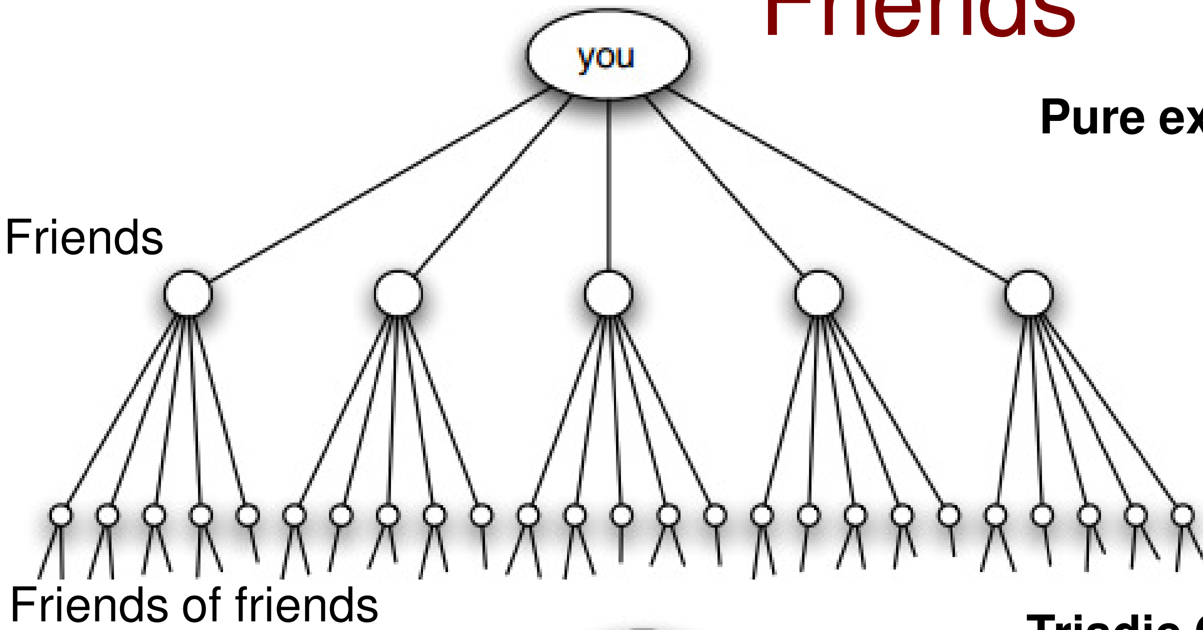
- Stanley Miligram (1960s) performed an experimental study to examine the average path length for social networks of people in the US.
- Miligram asked randomly chosen "starter" individuals to forward a letter to a designated "target" person in a suburb of Boston.
  - He provided the target's name, address, occupation and some personal info.
    - If the participant directly knew the target, they could mail it to them.
    - Otherwise, the participant should forward the letter to an acquaintance whom he knew and will be geographically more closer and likely to know the target. This is repeated until the letter reached the target.
- Roughly a third of the letters eventually arrived at the target.
  - The median path length took by these letters is SIX.
- The experiment demonstrated that several shorter paths exist in social networks and that people (without any sort of global map of the network) are effective at collectively finding the shorter paths, just based on their local connections and the geographic information on the target.

Lessons Learnt: Small-world Experiment

- Social networks tend to have very short paths between essentially arbitrary pairs of people.
  - The existence of these short paths has substantial consequences for the potential speed with which information, diseases, and other kinds of contagion can spread through society, as well as for the potential access that the social network provides to opportunities and to people with very different characteristics from one's own.

- Caveats:
  - The experiment does not clearly establish a statement quite as bold as "six degrees of separation between us and everyone else on this planet" – the paths were just to a single, fairly affluent target; many letters never got there; and attempts to recreate the experiment for people of lower status did not give desired results.
  - If we think of each person in the shorter path chain as the center of their own social world, then "six short steps" becomes "six worlds apart" – a change in perspective that makes six sound like a much larger number.

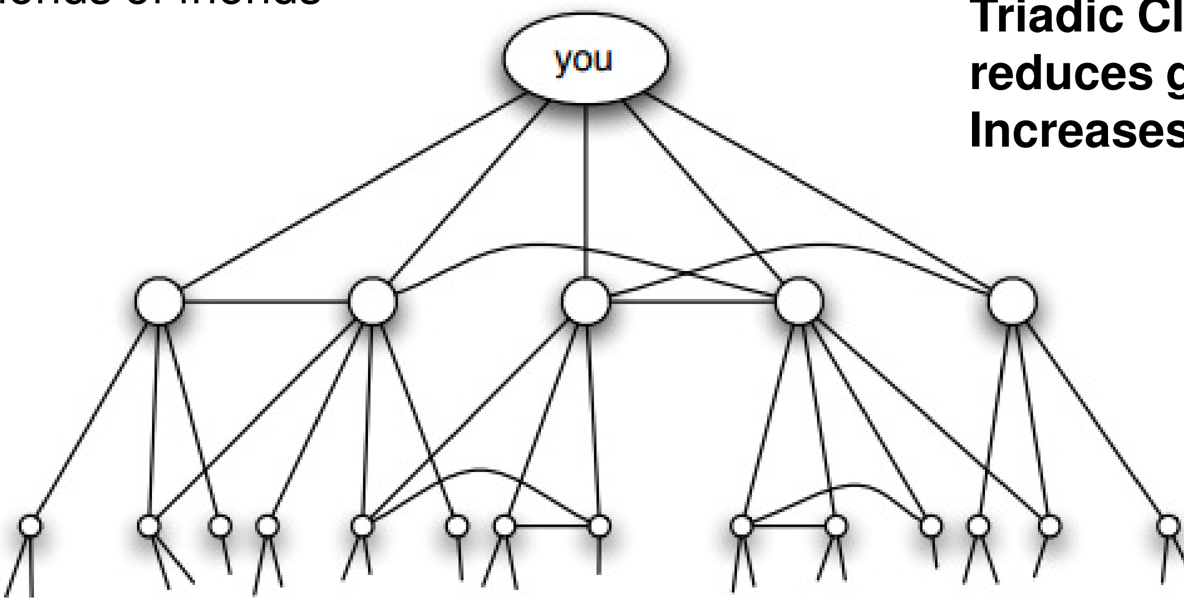
Triadic Closure and Growth Rate of New Friends



Pure exponential growth

No triadic closure

Zero clustering coefficient



Triadic Closure

reduces growth rate of new friends

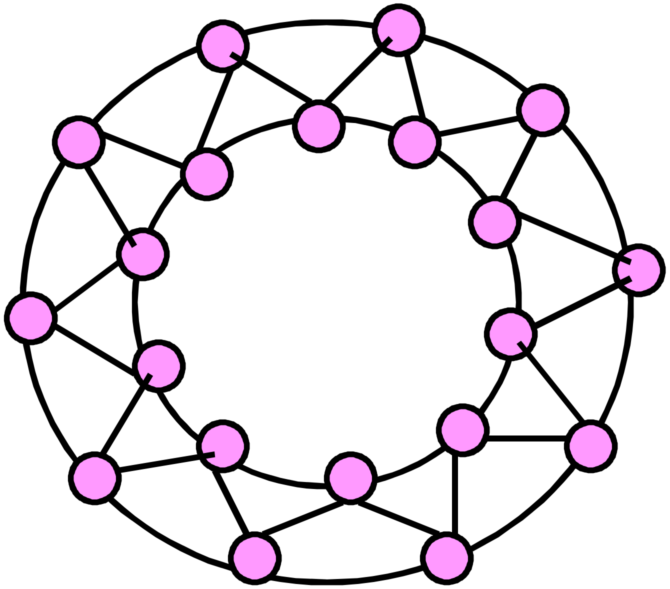
Increases clustering coefficient

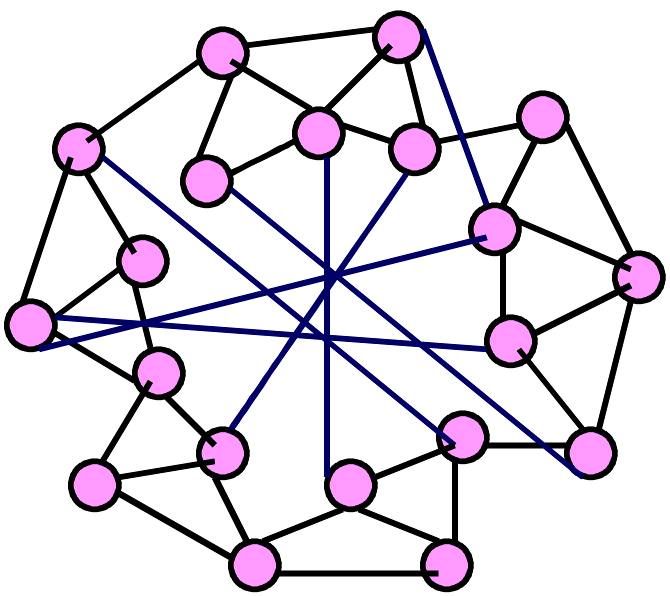
Modeling Small World Networks

- The ER model for random graphs provided shorter paths between any two nodes in the network. However, the ER graphs have a low clustering coefficient and triadic closures.
  - ER graphs have a constant, random and independent probability of two nodes being connected.

- The Watts and Strogatz model (WS model) accounts for clustering while retaining the short average path lengths of the ER model.
- The WS model interpolates between an ER graph and a regular ring lattice.

WS Model





- Watts and Strogatz (WS) Model: The WS model interpolates between an ER graph and a regular ring lattice.
  - Let N be the number of nodes and K (assumed to be even) be the mean degree.
  - Assume N >> K >> ln(N) >> 1.
  - There is a rewiring parameter β (0 ≤ β ≤ 1).
  - Initially, let there be a regular ring lattice of N nodes, with K neighbors (K/2 neighbors on each side).
  - For every node $n_i = n_0, n_1, \ldots, n_{N-1}$, rewire the edge (n_i, n_j), where $i < j$, with probability β. Rewiring is done by replacing (n_i, n_j) with (n_i, n_k) where n_k is chosen uniform-randomly among all possible nodes that avoid self-looping and link duplication.

β = 0 ➔ Regular ring lattice
β = 1 ➔ Random network

Small-World Network: WS Model

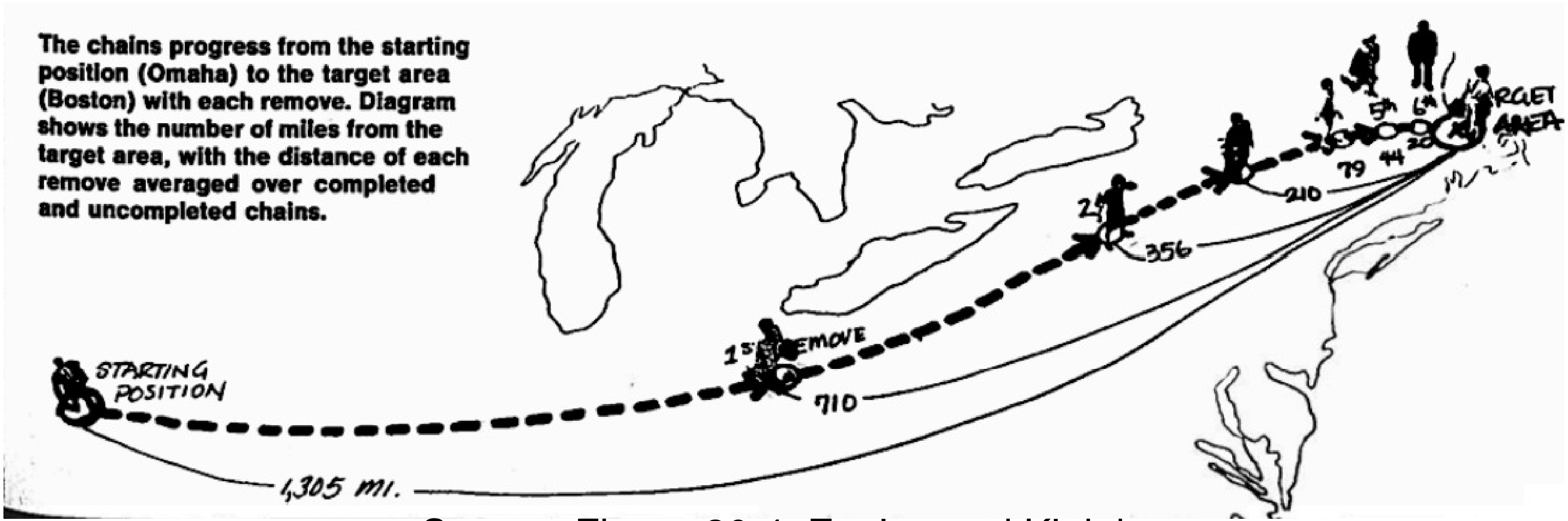
- The underlying lattice structure of the model produces a locally clustered network, and the random links dramatically reduce the average path lengths
- The algorithm introduces about ($\beta NK/2$) non-lattice edges.
- Average Path Length (β):
  - Ring lattice $L(0) = (N/2K) >> 1$
  - Random graph $L(1) = (\ln N / \ln K)$
  - For $0 < \beta < 1$, the average path length reduces significantly even for smaller values of β.
- Clustering Coefficient (β):

$$C(0) = \frac{3(K-2)}{4(K-1)} \qquad C(1) = \frac{K}{N} \qquad C'(\beta) = C(0) * (1-\beta)^3$$

  - For $0 < \beta < 1$, the clustering coefficient remains close to that of the regular lattice for low and moderate values of β and falls only at relatively high β.
- For low-moderate values of β, we thus capture the small-world phenomenon where the average path length falls rapidly, while the clustering coefficient remains fairly high.

Limitations of the WS Model

- The WS model introduced the notion of random edges to infuse shorter path lengths amidst larger clustering coefficient.
- However, the long-range edges span between any two nodes in the network and do not mimic the edges of different lengths seen in real-world networks (like in the US road map as in Milgram's experiment or airline map).
  - Path lengths could not be as small as they are in real networks.
  - Need some edges to nodes that are few hops away, rather than edges to some arbitrarily chosen nodes.
  - Cannot generate hubs as in scale-free networks.

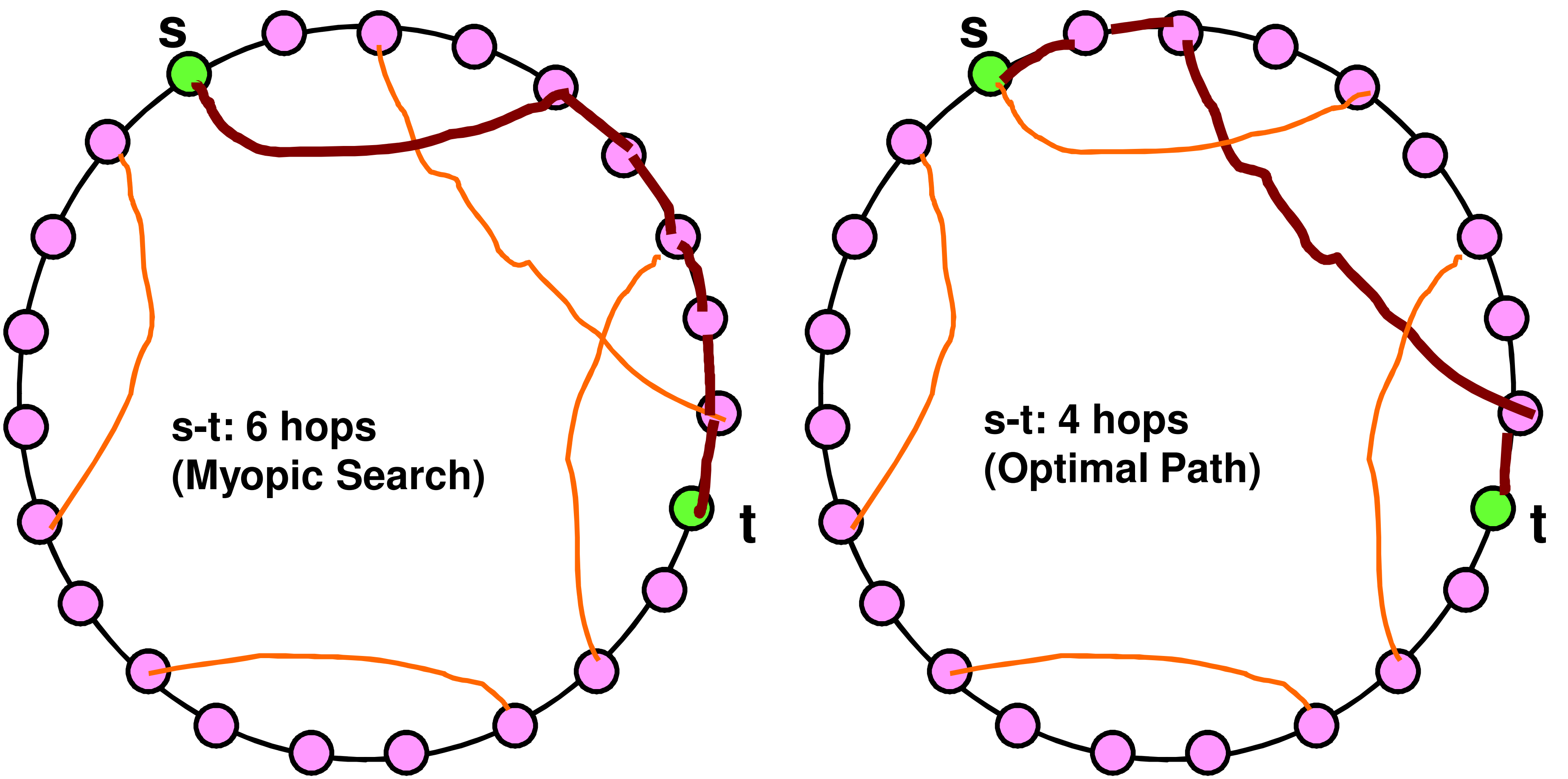


Source: Figure 20.4: Easley and Kleinberg

Enhancement to the WS Model

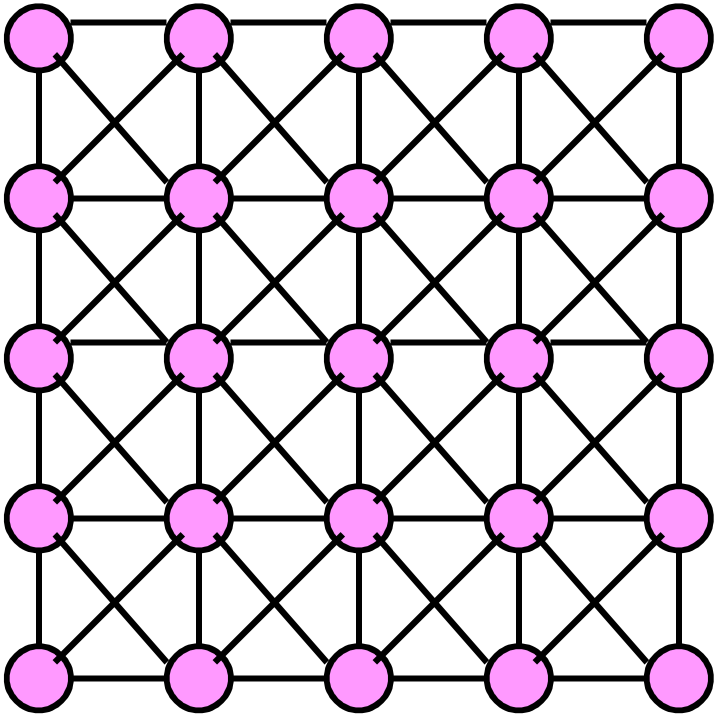
- In addition to the re-wiring parameter β, another parameter called the clustering exponent (q) is introduced.
- An (u, v) edge is selected for re-wiring with a probability β. After being selected, we do not randomly re-wire u with a node w. Instead, we pick a pair (u, w) for re-wiring with a probability of $[d(u, w)^{-q}] / 2\log n$, where
  - For optimal results, *q* must be the dimensionality of the network modeled. **For a ring lattice, q = 1.**
  - *n* is the number of nodes in the network.
  - d(u, w) is the minimum number of hops between u and w in the original network layout (before enhancement)
    - The ring lattice is a single-dimension network
    - A grid is a two-dimensional network.
  - To implement this enhancement in simulations, we generate a random number between 0 to 1; the (u, w) pair whose $[d(u, w)^{-dim}] / 2\log n$ value is closest and above the random number generated is chosen for re-wiring.
- With this re-wiring model, if routed optimally, (on average) the # hops in the path to the target is expected to reduce by a factor of 2 with every additional hop in the path [log*n* hops]

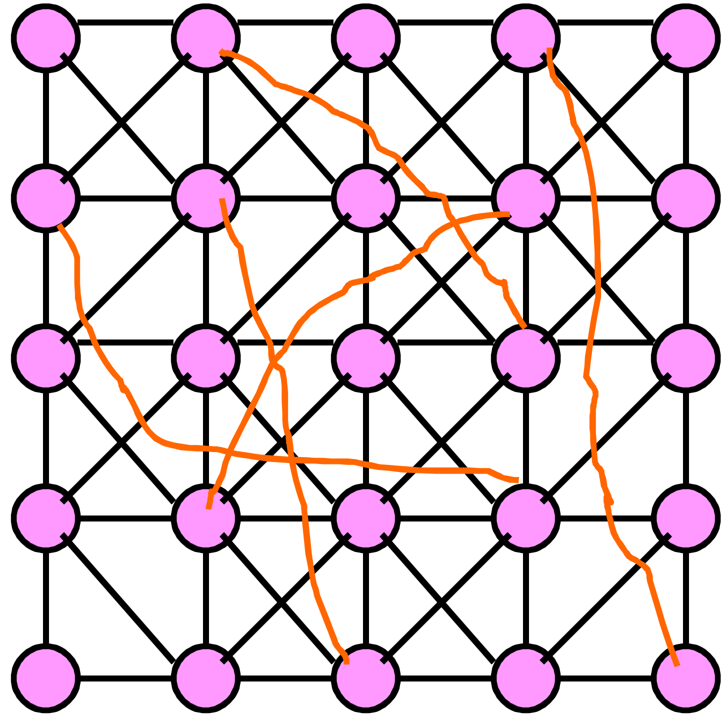
Myopic Search



Myopic Search: The source node forwards the packet to the neighbor node that is geographically closer to the target (under the underlying ring lattice). The source node does not know the long-range contacts of the subsequent nodes. This could lead to sub optimal paths; but still fewer hops than in a ring lattice.

WS Model for Grid Lattice (2-dim)



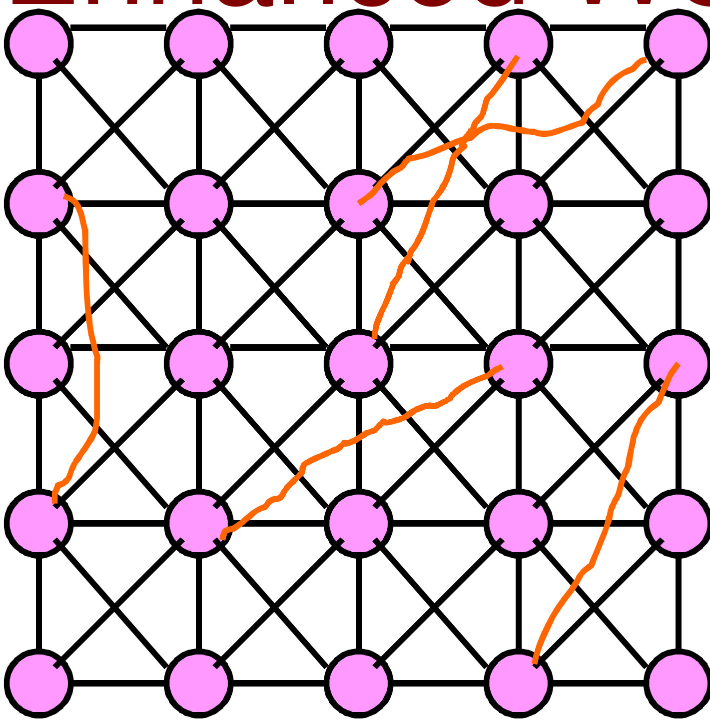


We have a nxn two-dimensional grid.
The nodes are identified with lattice points i.e., a node v is identified with the lattice point (i, j) with i, j = {1, 2, …, n}

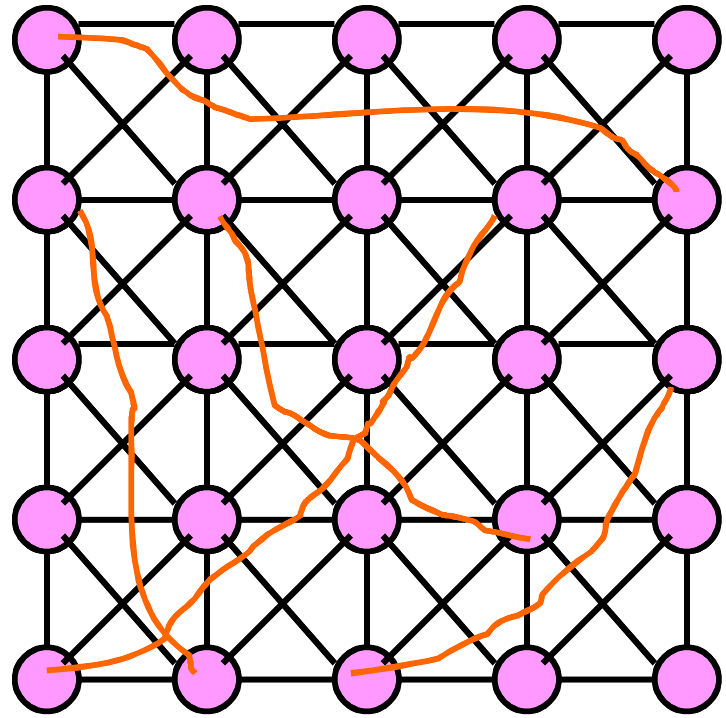
Distance between two nodes at (i, j) and (k, l) is: $| i - k | + | j - l |$

For every node u, we remove one of its associated edges to the neighbor nodes with a probability β and connect the node to a randomly chosen node v, as long as there is no self-loops and link duplication.

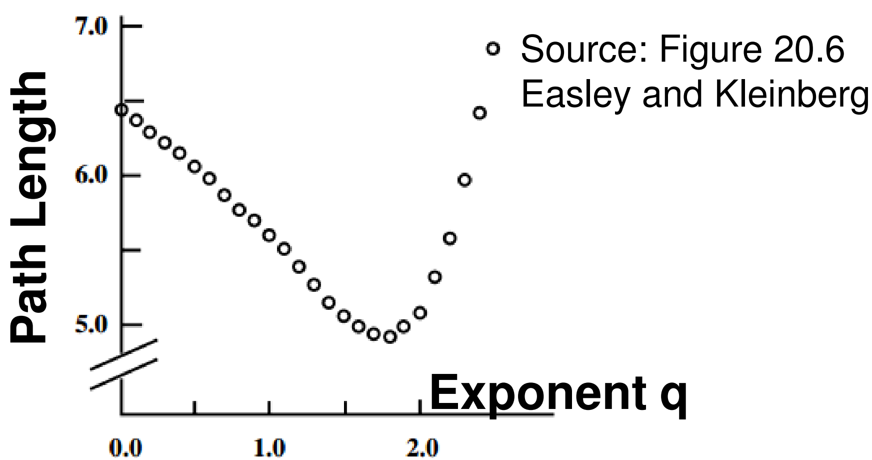
Enhanced WS Model: Grid Lattice



High Clustering Exponent Value



Low Clustering Exponent Value



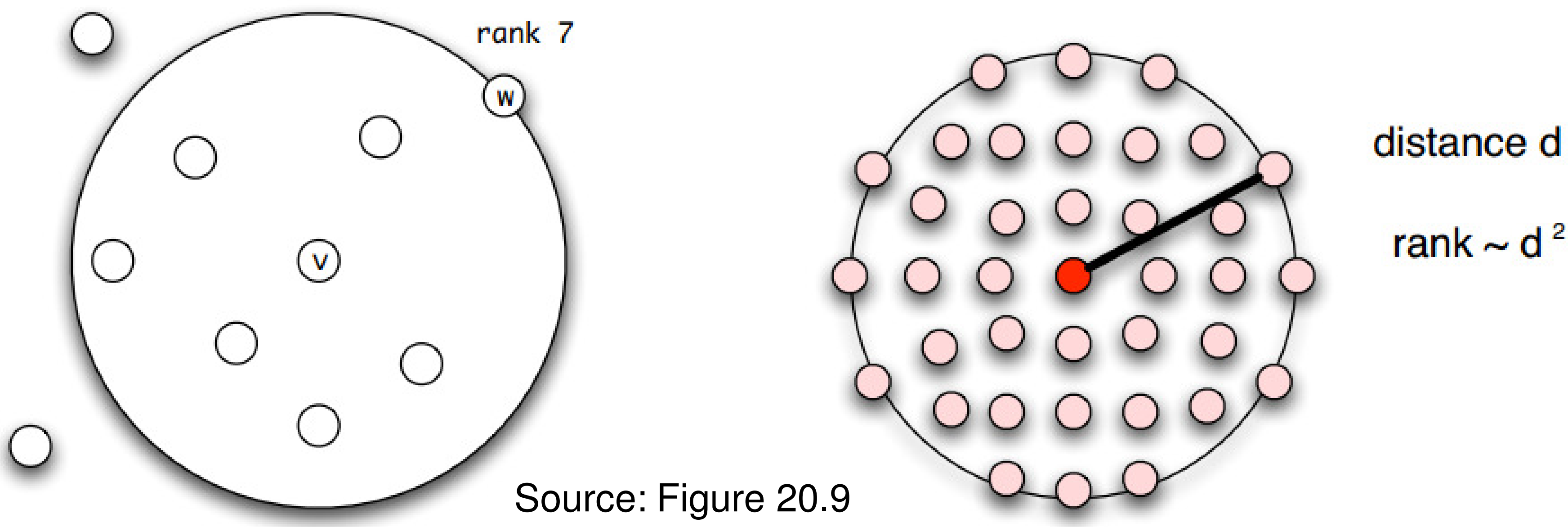
Probability of a long-range edge (u, v)

Optimal q = 2

$$\frac{d(u,v)^{-q}}{2\log n}$$

Rank-based Long-range Weak Ties

- For non-uniform networks in which distances between nodes do not follow a uniform distribution
  - For every node v: we rank the nodes based on their distance from v.
  - We replace the [distance(v, w)]$^{-q}$ with rank^{-1} in the probability for the two nodes (v, w) to have a long-range weak tie.

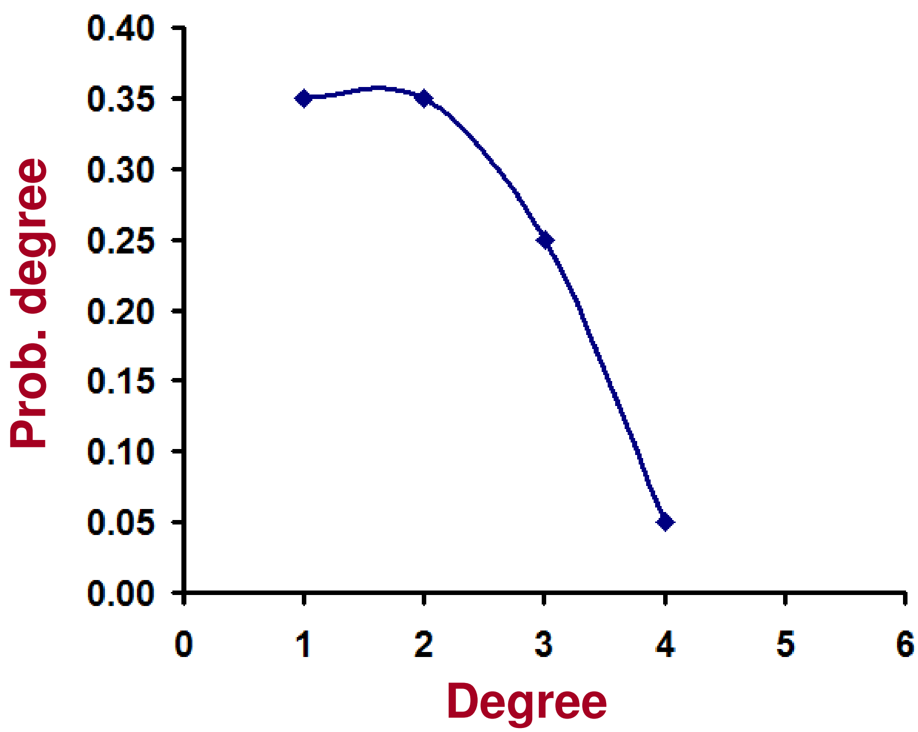


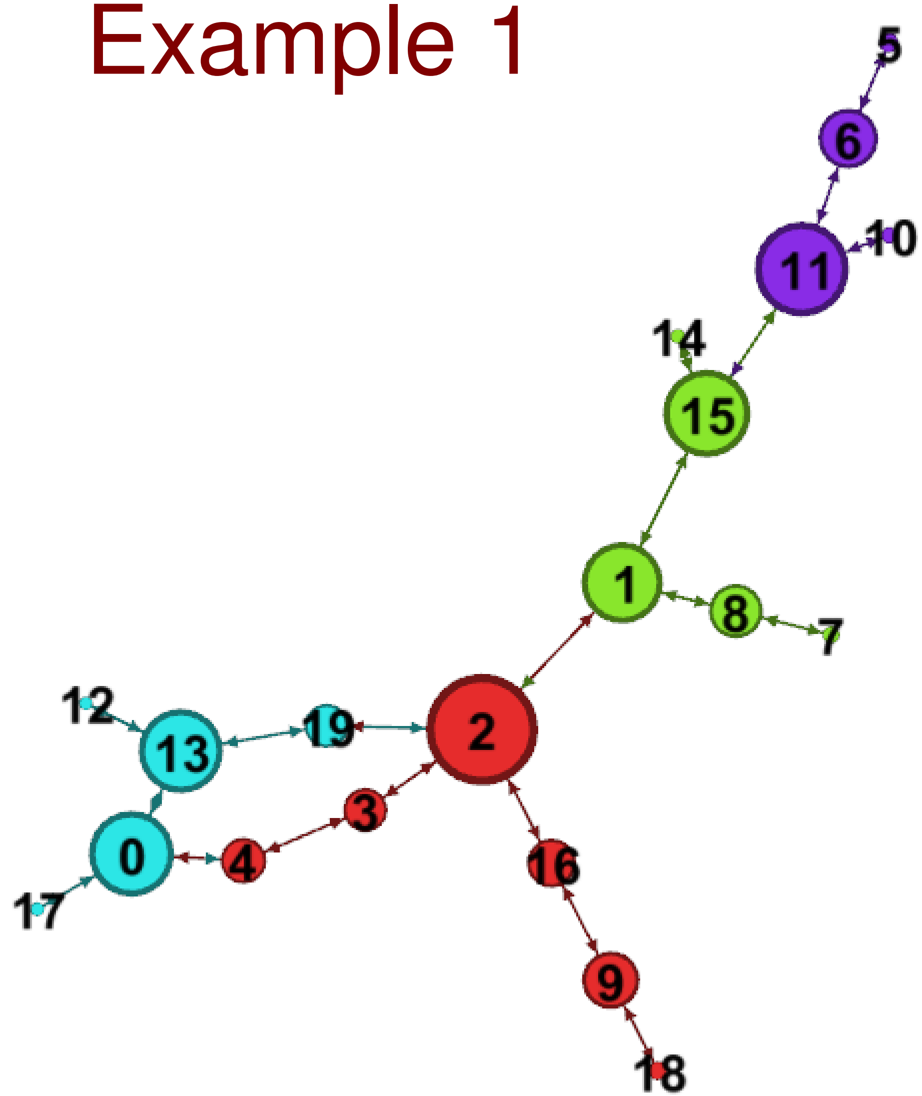
Source: Figure 20.9
Easley and Kleinberg

CINET: SW Model Example 1

Total # nodes: 20
K-neighbors per node (ring): K = 4
Probability of re-wiring p = 0.7

Avg. Node degree: 2.0
Network diameter: 9.0
Avg. clustering coeff: 0
Communities: 4 (modularity score: 0.54)
Avg. Path Length = 3.963



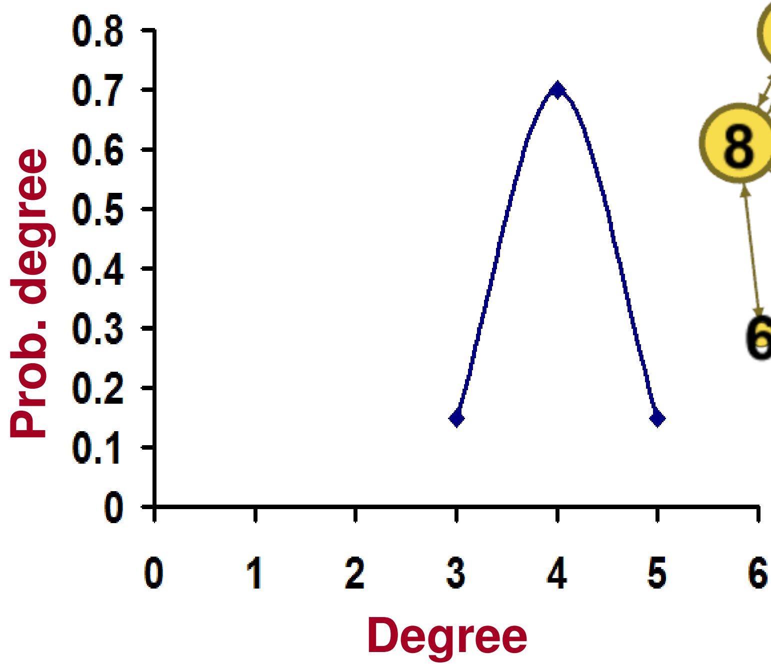


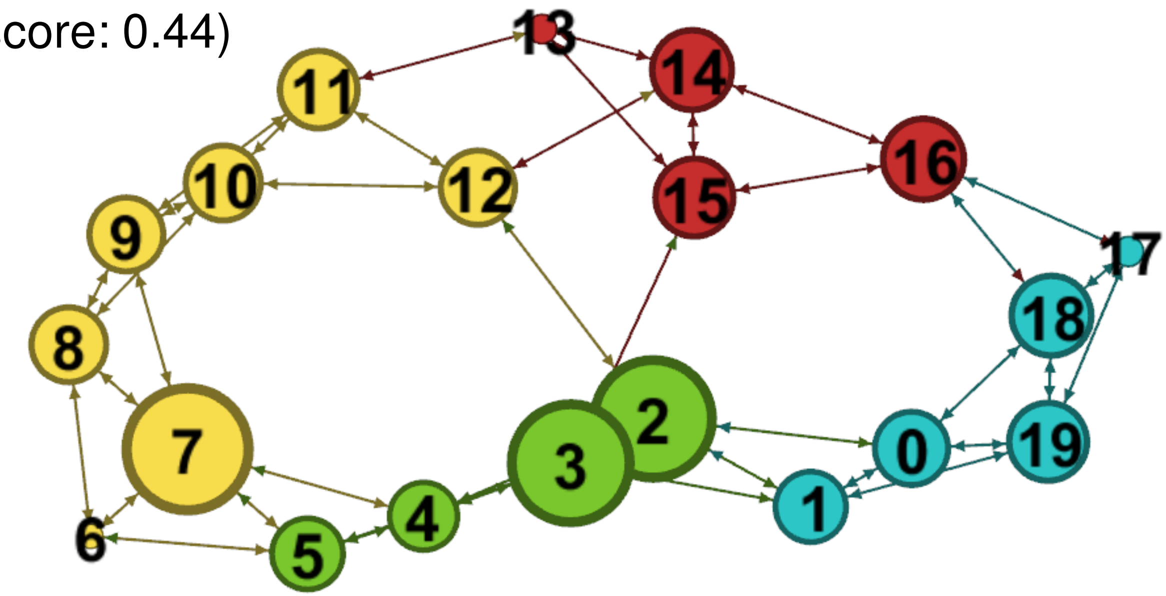
Pagerank of the vertices (size)
Hubs – have larger Pagerank

CINET: SW Model Example 2

Total # nodes: 20
K-neighbors per node (ring): K = 4
Probability of re-wiring p = 0.2

Avg. Node degree: 4.0
Network diameter: 4.0
Avg. clustering coeff: 0.435
Communities: 4 (modularity score: 0.44)
Avg. Path Length = 2.526





Pagerank of the vertices (size)
Hubs – have larger Pagerank