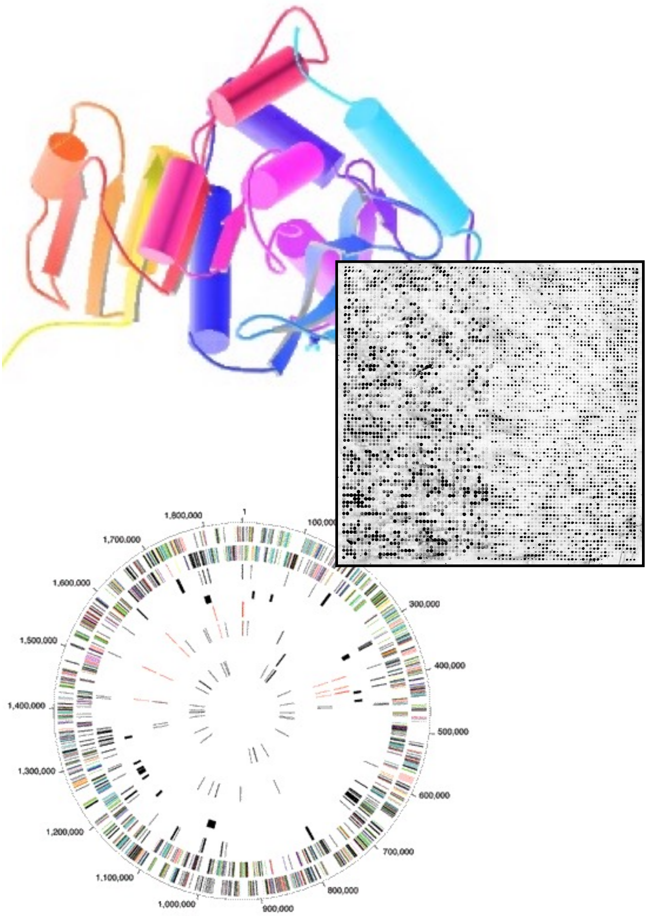
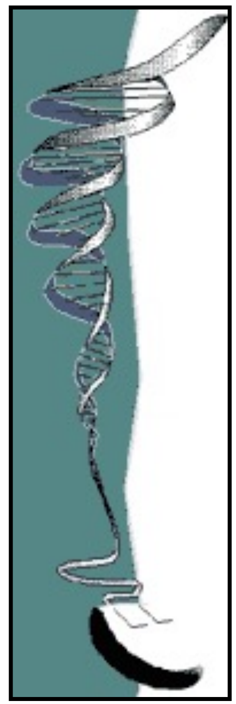
Biomedical Data Science:

Analysis of Network Topology C – Network Generation Models





Mark Gerstein, Yale University

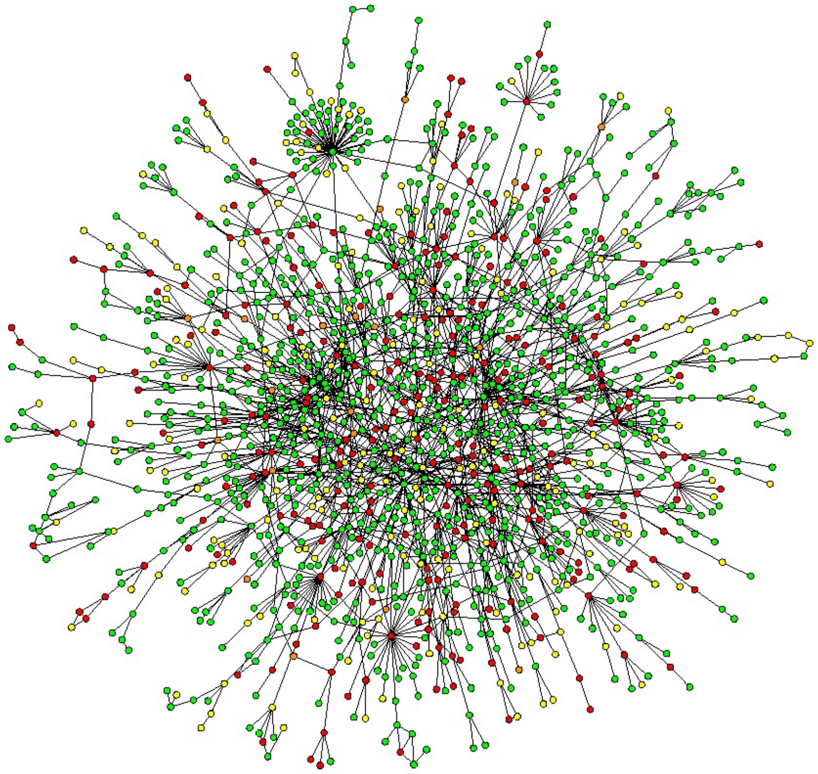
gersteinlab.org/courses/452

(last edit in spring ‘21)

Network Topology

Simple Mathematical Models for Interpreting Complex Topology: ER Model & Small World Networks

Models for networks of complex topology



- Erdos-Renyi (1960)
- Watts-Strogatz (1998)
- Barabasi-Albert (1999)

A Barabási & R Albert "Emergence of scaling in random networks," ***Science*** **286, 509-512 (1999).**

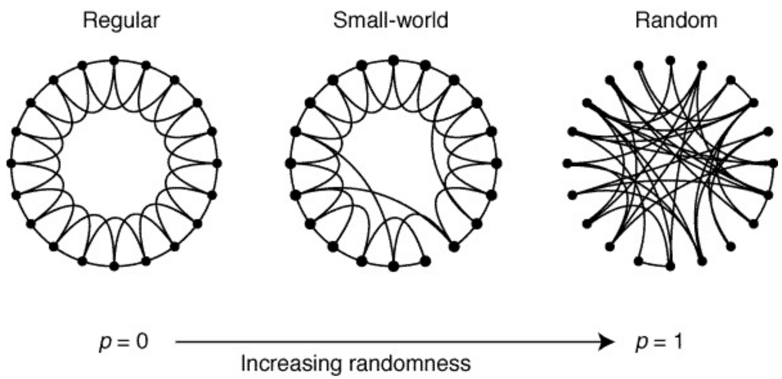
The Erdős-Rényi [ER] model (1960)

- Start with N vertices and no edges
- Connect each pair of vertices with probability P_{ER}

Important result: many properties in these graphs appear quite suddenly, at a threshold value of $P_{ER}(N)$

-If $P_{ER} \sim c/N$ with $c<1$, then almost all vertices belong to isolated trees
-Cycles of all orders appear at $P_{ER} \sim 1/N$

The Watts-Strogatz [WS] model (1998)



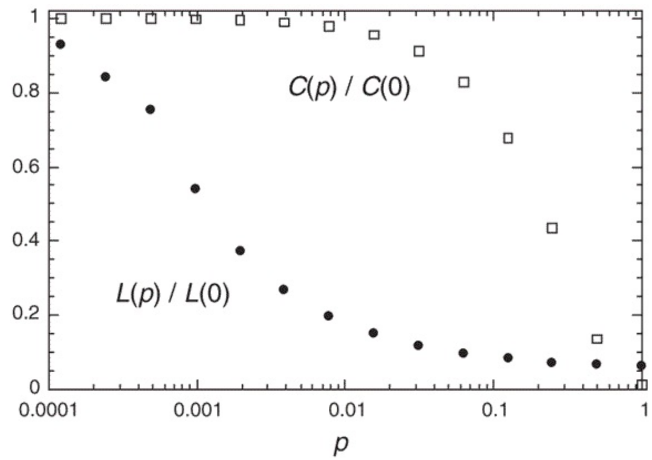
- Start with a regular network with N vertices
- Rewire each edge with probability p

For p=0 (Regular Networks):
- high clustering coefficient
- high characteristic path length

For p=1 (Random Networks):
- low clustering coefficient
- low characteristic path length

QUESTION: What happens for intermediate values of p?

1) There is a broad interval of p for which L is small but C remains large



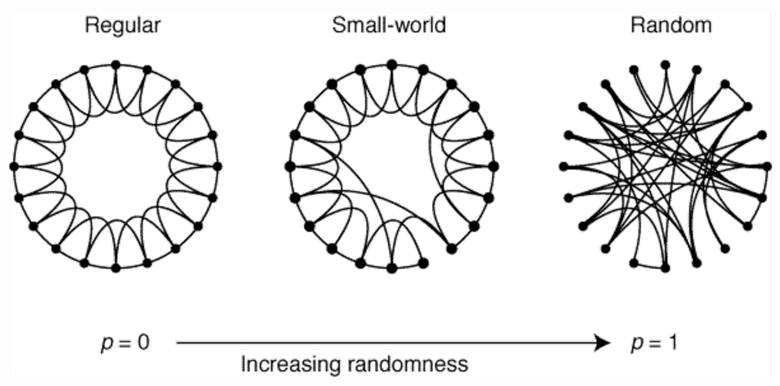
2) Small world networks are common :

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Small world network

- A simple connected graph G exhibiting two properties:
  - **Large Clustering Coefficient:** Each vertex of G is linked to a relatively well-connected set of neighboring vertices, resulting in a large value for the clustering coefficient C(G);
  - **Small Characteristic Path Length**: The presence of short-cut connections between some vertices results in a small characteristic path length L(G).



- local connectivity and global reach

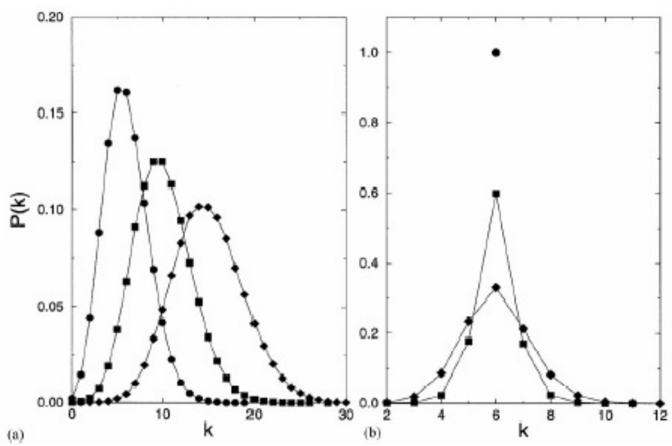
Watts and Strogatz (1998), Nature, Collective dynamics of 'small-world' networks

Network Topology

Simple Mathematical Models for Interpreting Complex Topology: BA Model & Scale Free Networks

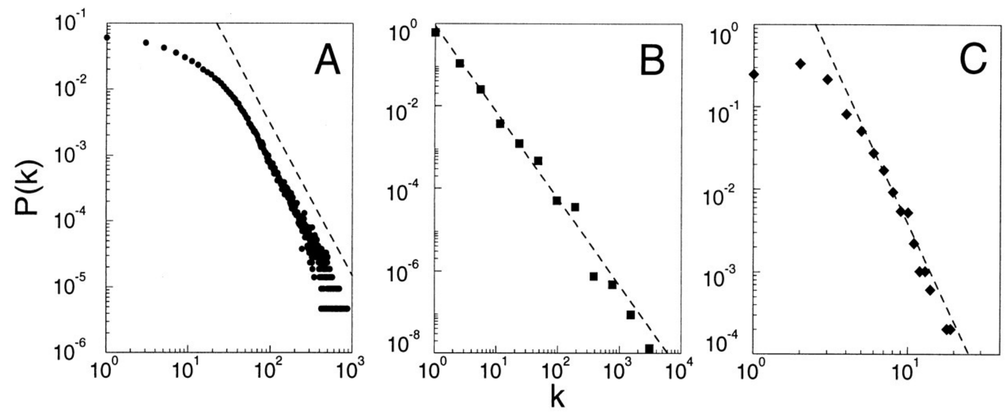
The Barabási-Albert [BA] model (1999)

Look at the distribution of degrees



ER Model

WS Model



actors

power grid

www

The probability of finding a highly connected node decreases exponentially with k

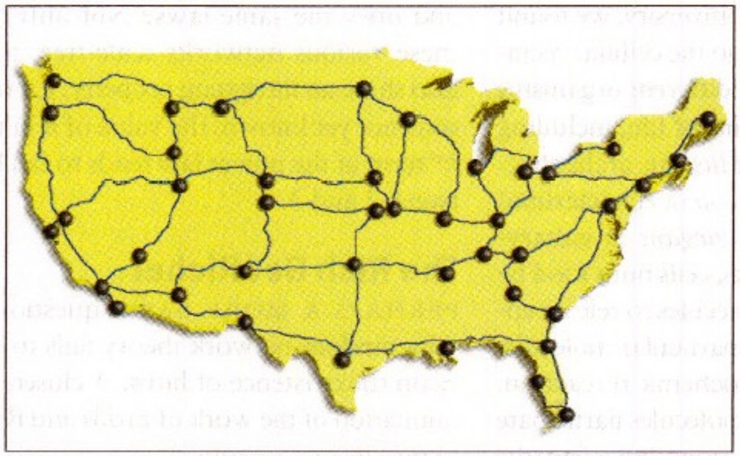
$$P(K) \sim K^{-\gamma}$$

Random v Scale-free Networks

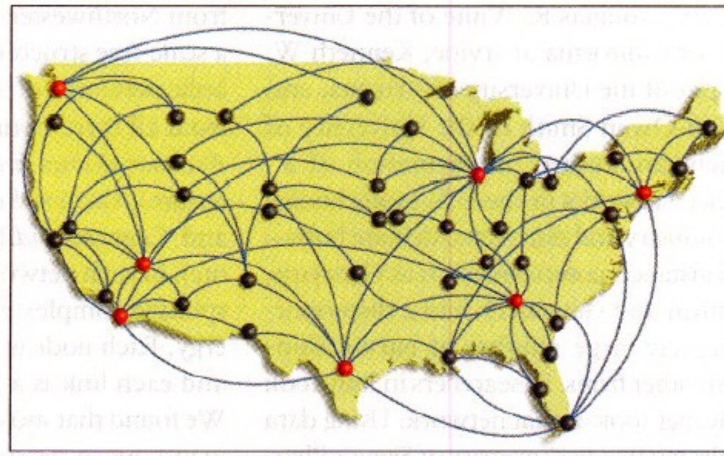
RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)—nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

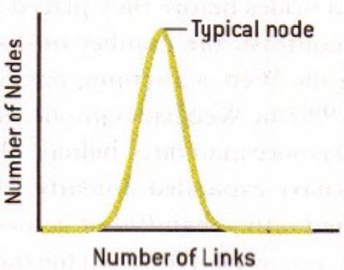
Random Network



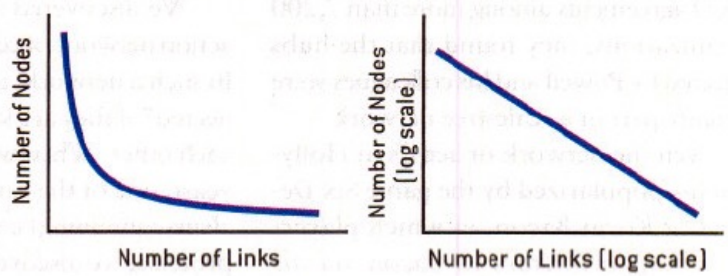
Scale-Free Network



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages



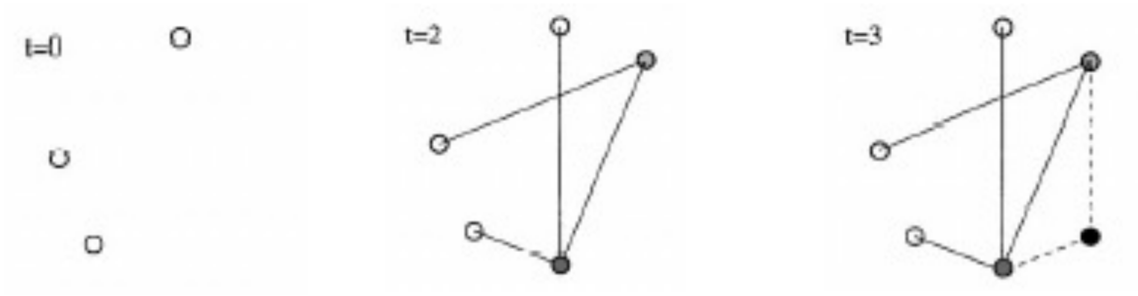
[From Barabasi & Bonabeau, *Sci. Am.*, May '03]

- two problems with the previous models:
    1. N does not vary
    2. the probability that two vertices are connected is uniform

• GROWTH: starting with a small number of vertices m_0 at every timestep add a new vertex with $m \leq m_0$

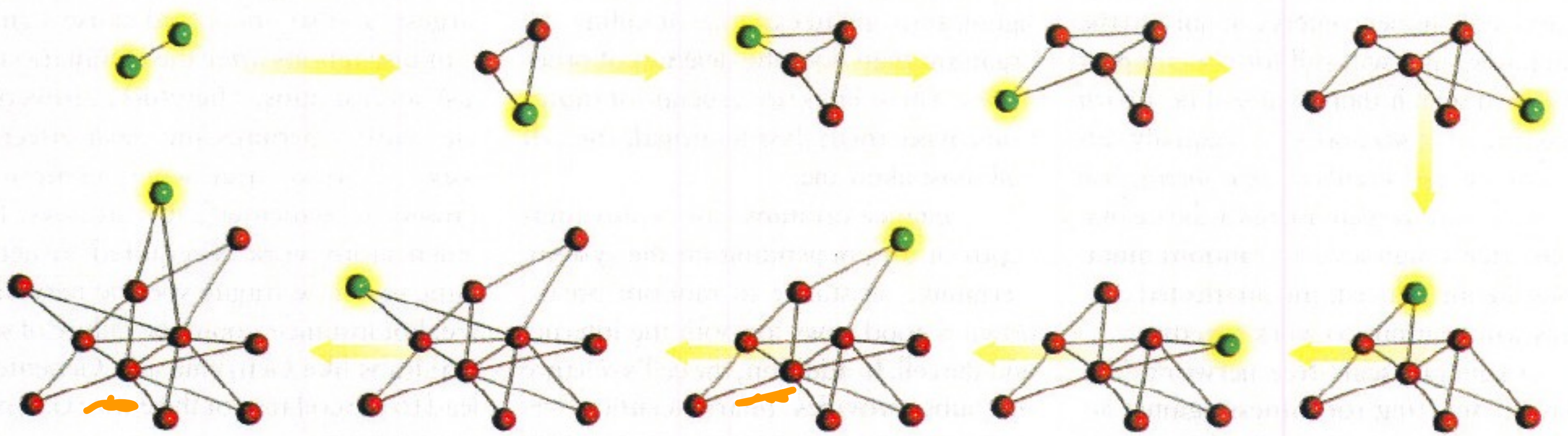
• PREFERENTIAL ATTACHMENT: the probability Π that a new vertex will be connected to vertex i depends on the connectivity of that vertex:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$



Birth of Scale-Free Network

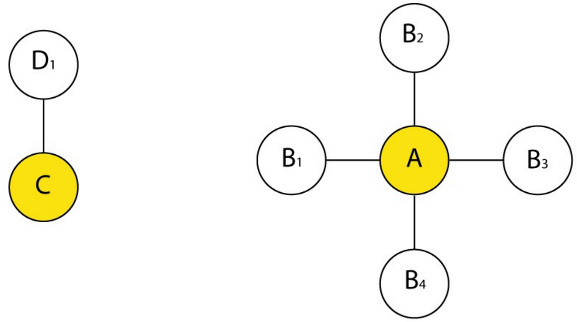
A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



[From Barabasi & Bonabeau, *Sci. Am.*, May '03]

SCALE FREENESS GENERALLY EVOLVES THROUGH PREFERENTIAL ATTACHMENT (THE RICH GET RICHER)

The Duplication Mutation Model



Gene duplication

The interaction partners of A are more likely to be duplicated

Description

- Theoretical work shows that a mechanism of preferential attachment leads to a scale-free topology

  ("The rich get richer")

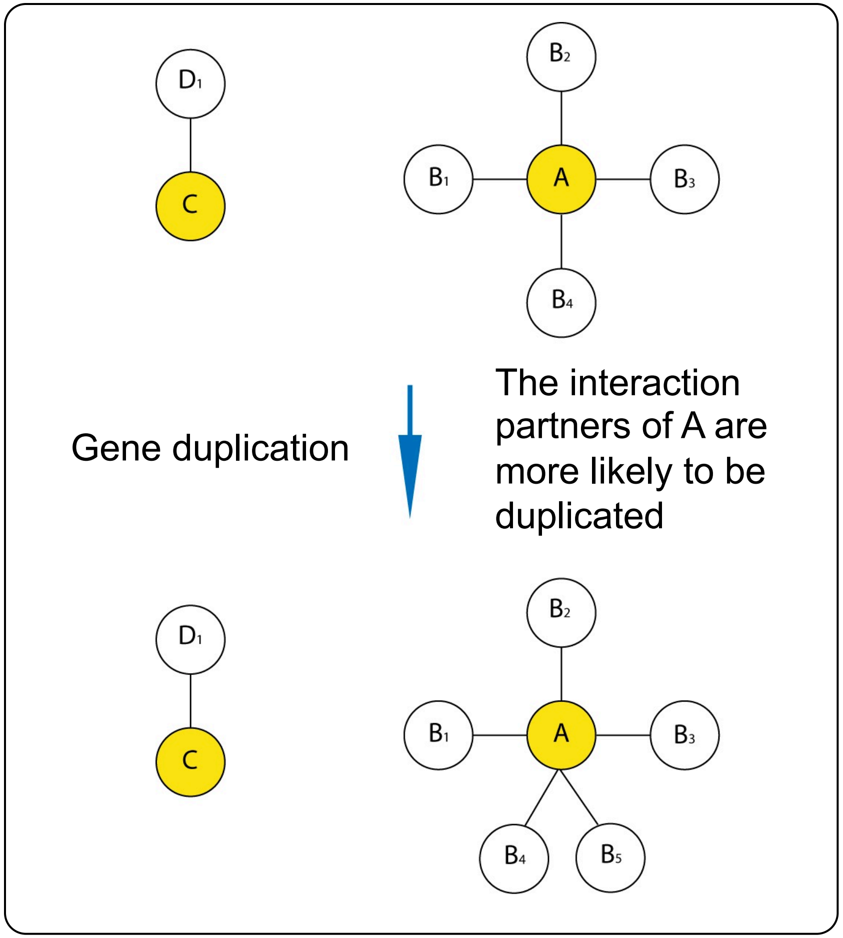
- In interaction network, gene duplication followed by mutation of the duplicated gene is generally thought to lead to preferential attachment

- Simple reasoning: The partners of a hub are more likely to be duplicated than the partners of a non-hub

Source: Albert et al. *Rev. Mod. Phys.* (2002) and Middendorf et al. *PNAS* (2005)

SCALE FREENESS GENERALLY EVOLVES THROUGH PREFERENTIAL ATTACHMENT (THE RICH GET RICHER)

The Duplication Mutation Model



Description

- Theoretical work shows that a mechanism of preferential attachment leads to a scale-free topology

  ("The rich get richer")

- In interaction network, gene duplication followed by mutation of the duplicated gene is generally thought to lead to preferential attachment

- Simple reasoning: The partners of a hub are more likely to be duplicated than the partners of a non-hub

Source: Albert et al. *Rev. Mod. Phys.* (2002) and Middendorf et al. *PNAS* (2005)