Network Science: Clustered models for network formation

Ozalp Babaoglu
Dipartimento di Informatica — Scienza e Ingegneria
Università di Bologna
www.cs.unibo.it/babaoglu/

Clustered models

- How to extend the ER model to be a better predictor of real network properties
- First, address the poor prediction of clustering
- ER model ignorant of current network structure in adding edges — all edges have exactly the same probability of appearing in the network regardless of their position
- In real networks, the formation of edges is often highly biased
- Bias towards connecting friends of friends

Clustered models

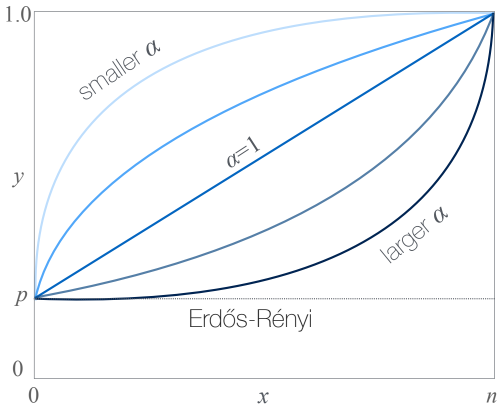
- In social networks, people introduce their friends to each other
- People that have common friends have more occasions to meet each other and become friends themselves — ***triadic closure***
- People who have common friends often also have common interests — ***homophily***
- First idea: select edges randomly, but with a bias towards friends of friends
- The more common neighbors two nodes share, the more likely they will be connected

Clustered models
The α model

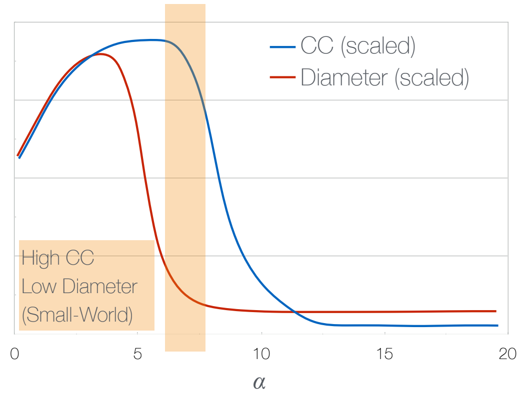
- Bias the connections towards nodes that have common neighbors
- For some arbitrary pair of nodes, let x denote the number of neighbors they currently have in common
- Let y be the probability of adding an edge between a pair of nodes that have x common neighbors
- Assume that $y \sim p + (x/n)^\alpha$ for some constants p and α

Clustered models
The α model

$$y \sim p + \left(\frac{x}{n}\right)^{\alpha}$$



Clustered models
The α model



Clustered models
The α model

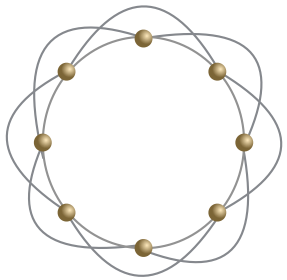
- The model needs to be “tuned” (setting α) and is able to achieve high clustering
- Get small diameter “for free”

Clustered models
Watts-Strogatz

- Start with a highly regular network capturing relations that correspond to geographic, social proximity
- Such a network typically has high clustering but large diameter
- Idea (“rewire”): replace (a few) local edges with random “long distance” (short-cut) edges corresponding to occasional contacts outside of usual social circles
- A balanced set of “local” and “long distance” edges may exhibit high clustering while reducing the diameter

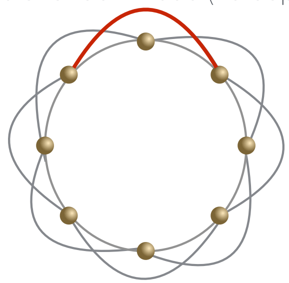
Clustered models
Watts-Strogatz

- Start with a K-regular lattice (ring, grid, cube, etc.) where each node is connected to its K nearest neighbors
- Example: n=8, K=4



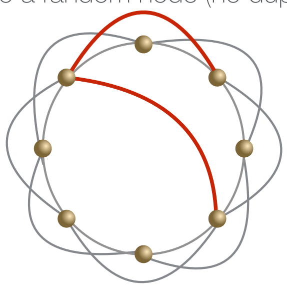
Clustered models
Watts-Strogatz

- Clustering coefficient is 0.5 while the diameter is order n (need to go half-way around the ring)
- "Rewire" with probability q to a random node (no duplicates, no self edges)



Clustered models
Watts-Strogatz

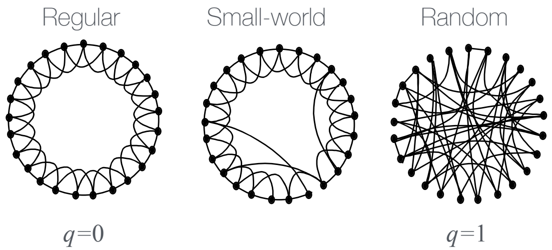
- Clustering coefficient is 0.5 while the diameter is order n (need to go half-way around the ring)
- "Rewire" with probability q to a random node (no duplicates, no self edges)



- NetLogo *Library/Networks/Small Worlds* and "SmallWorldWS"

Clustered models
Watts-Strogatz

- As the rewiring probability q increases from 0 to 1, we get more and more random networks passing through a region called "small-world" networks

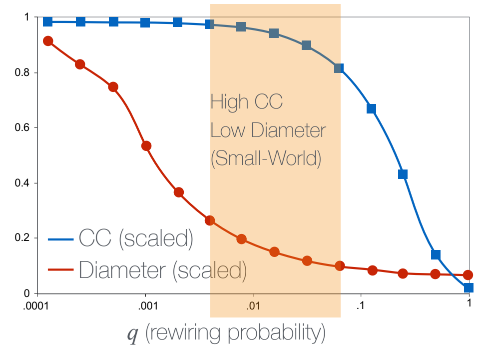


Clustered models
Watts-Strogatz

- Origin of small-world networks
  - Diameter is governed by the *number* of random shortcuts (qn)
  - Clustering is governed by the *fraction* of random shortcuts (q)
- *Fact*: roughly 5 random shortcuts reduce average path length by factor of 2, independent of n
- For large n, a small number of random shortcuts will reduce the diameter substantially while leaving clustering mostly unchanged

Clustered models
Watts-Strogatz

- Just like the α model, Watts-Strogatz needs to be "tuned" by setting the rewire probability q



Clustered models
Watts-Strogatz

- Clustering coefficient of K-regular lattice is $\frac{3(K-2)}{4(K-1)}$
- Converges to 3/4 in the limit for large K
- Average path length for a d-dimensional hypercube scales as $n^{1/d}$ which grows much faster than logarithmic
- For the WS model, the clustering coefficient is

$$\frac{3K(K-1)}{2K(2K-1) + 8qK^2 + 4q^2K^2}$$

- While the average path length is

$$\frac{n^{1/d}}{K} f(qKn)$$

where $f(qKn)$ is a universal scaling function