Biological Networks

Lectures 8-9 : February 11, 2010

Network Models
(Random Graphs)

Network Models

We will cover the following network models:

I. Erdös–Rényi random graphs
II. Generalised random graphs (with the same degree distribution as the data networks)
III. Small-world networks
IV. Scale-free networks
   1) preferential attachment networks (growth model)
   2) gene duplication and mutation networks
V. Hierarchical model
VI. Geometric random graphs
VII. Stickiness index-based network model

I. Erdös–Rényi random graphs (ER)

We are trying to model a real-world network $G(V,E)$ with $|V|=n$ and $|E|=m$.

An ER graph that models it is constructed as follows:

- It has n nodes
- edges are added between pairs of nodes uniformly at random with the same probability p
- there are two equivalent methods for constructing ER graphs:
  - pick p so that the resulting model network has m edges. This model is denoted by $G_{n,p}$
  - pick randomly m pairs of nodes and add edges between them with probability 1. This model is denoted by $G_{n,m}$.

I. Erdős–Rényi random graphs (ER)

Number of edges, |E|, in $G_{n,p}$ is:

- $|E| = \binom{n}{2} p$ = pn(n-1)/2

- average degree:

$$z = \frac{2|E|}{n} = \frac{2\binom{n}{2}p}{n} = (n-1)p$$

I. Erdős–Rényi random graphs (ER)

Many properties of ER can be proven theoretically

(See Bollobas, "Random Graphs," 2002)

Examples:

- when $m=n/2$, suddenly the giant component emerges, i.e.,
  - one connected component of the network (connected component) has $O(n)$ nodes
  - the next largest connected component has $O(\log(n))$ nodes

I. Erdős–Rényi random graphs (ER)

- The **degree distribution** is Binomial:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

For large *n*, this can be approximated by the Poisson distribution:

$$P(k) = \frac{z^k e^{-z}}{k!}$$

where z is the average degree

I. Erdős–Rényi random graphs (ER)

- Currently available ***biological networks*** have *power-law degree distribution*.
- Thus, ER is not a good model for biological networks with respect to degree distribution

I. Erdős–Rényi random graphs (ER)

- **Clustering coefficient**, *C*, of ER is low.
- *C=p,* since probability *p* of connecting any two nodes in an ER graph is the same, regardless of whether the nodes are neighbours.
- Thus, ER is not a good model for biological networks with respect to the clustering coefficient, since ***biological networks*** have *high clustering coefficients*.

I. Erdős–Rényi random graphs (ER)

- **Average diameter** of ER graphs is small, it is equal to $\frac{\ln(n)}{\ln(z)}$
- This property of ER networks models well the average diameters of biological networks, since ***biological networks*** have *small average diameters*

I. Erdős–Rényi random graphs (ER)

Summary:

	Deg. Distr.	Clust. Coef.	Avg. Diam.
Real Networks	Power-law	High	Small
ER	Poison	Low	Small

Since only one property of ER models the data well, better fitting models are sought.

II. Generalized random graphs

- preserve the degree distribution of data ("ER-DD")
- They are constructed as follows:
  - an ER-DD network has *n* nodes (this is the number of nodes of the data)
  - edges are added between pairs of nodes using the "stubs method" as follows:

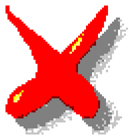
II. Generalized random graphs

- The “stubs method” for constructing ER-DD graphs:
  - the number of “stubs” (to be filled by edges) is assigned to each node in the model network according to the degree distribution of the real network to be modelled.
  - edges are created between pairs of nodes with stubs picked at random.
  - after an edge is created, the number of stubs left available at the corresponding “end nodes” of the edges is decreased by one.
  - we do not allow multiple edges between the same pair of nodes.

II. Generalized random graphs

Summary:

	Deg. Dist.	Clust. Coef.	Avg.Diam.
Real Networks	Power-law	High	Small
ER-DD	Power-law	Low	Small

- So, two global network properties of biological networks are matched by ER-DD.
- How about local network properties?

II. Generalized random graphs

Local network properties of ER and ER-DD:

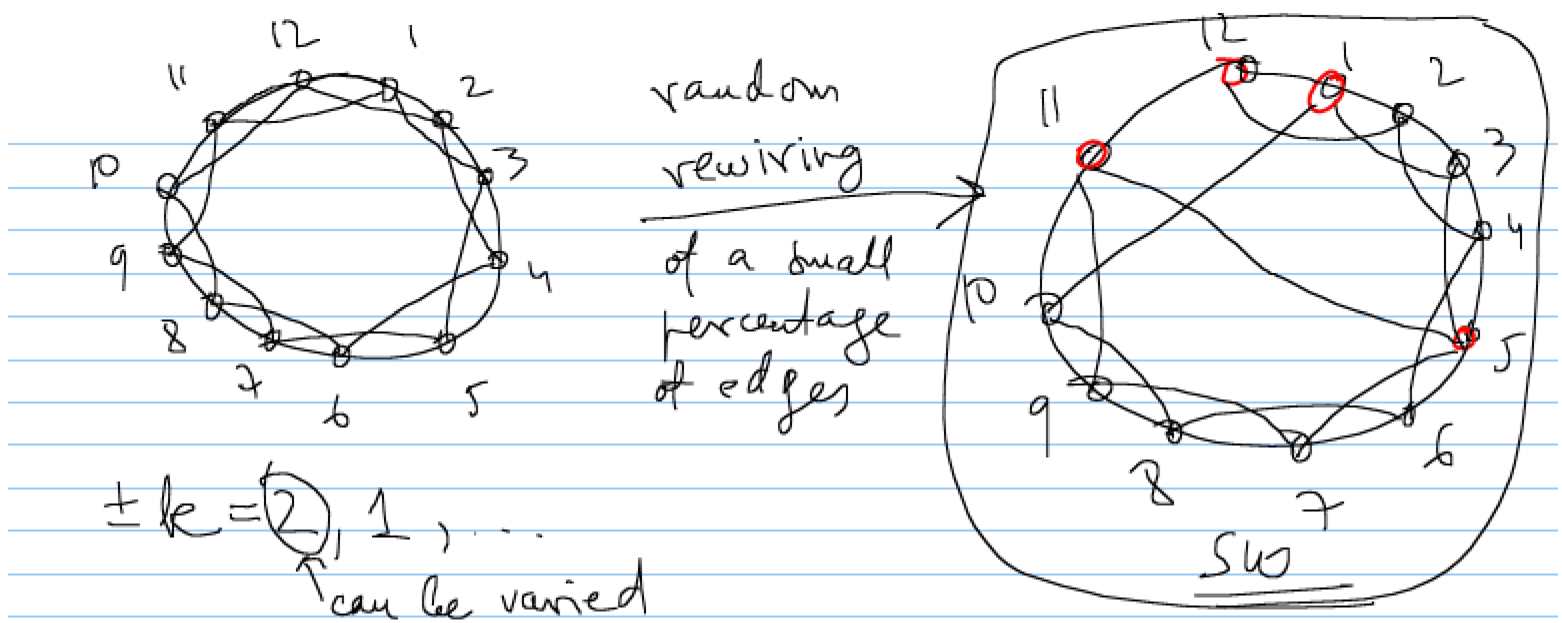
- Graphlet frequencies:
    - low-density graphlets are over-represented in ER and ER-DD
    - data have lots of dense graphlets, since they have high clustering coefficients

III. Small World Networks ("SW")

(Watts and Strogatz, 1998)

- Created from regular ring lattices by random rewiring of a small percentage of their edges

E.g.



III. Small World Networks ("SW")

SW networks have:

- high clustering coefficients - introduced by "ring regularity"
- However, regular ring lattices have large average diameters, BUT:
  - this can be solved by randomly re-wiring a small percentage of links

III. Small World Networks ("SW")

Summary:

	Deg. Dist.	Clust. Coef.	Avg.Diam.
Real Networks	Power-law	High	Small
SW	Poisson?	High	Small
	✗	✓	✓