Key Concepts & Measures in SNA

Global Structure of Networks

- A Social network can be represented as a graph *G = (V,E)*

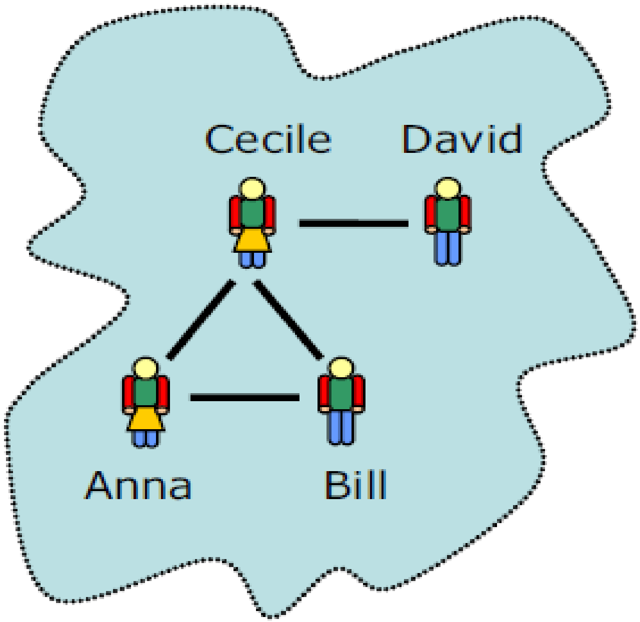
- V - finite set of vertices
- E - finite set of edges such that $E \subseteq V \times V$

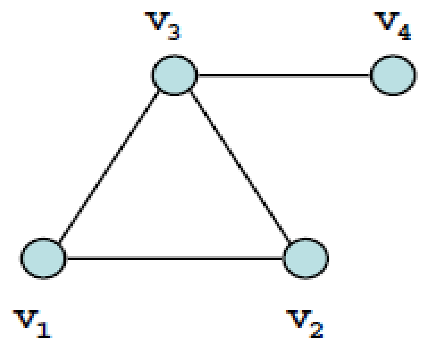
$$M := (m_{i,j})_{n*n} \text{ where } n = |V|, m_{i,j} = \left\{ \begin{array}{c|l} 1 & (v_i, v_j) \in E \\ 0 & otherwise \end{array} \right\}$$

- For weighted (valued) graphs, its characteristic matrix is defined as

$$m_{i,j} = \left\{ \begin{array}{l|l} w(e) & (v_i, v_j) \in E \\ 0 & otherwise \end{array} \right\}$$

Social Network represented as Graph

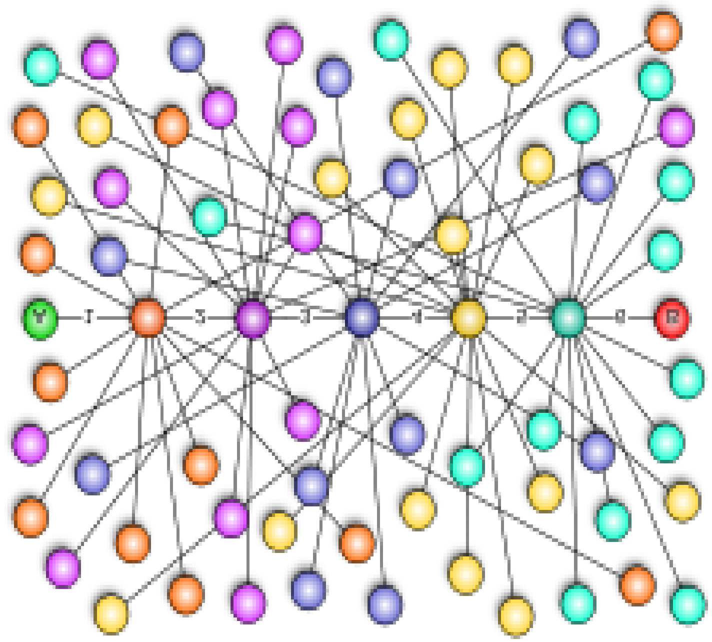




	v_1	v_2	v_3	v_4
v_1	0	1	1	0
v_2	1	0	1	0
v_3	1	1	0	1
v_4	0	0	1	0

Contd…

- Major breakthrough in SNA - American psychologist Stanley Milgram proposed the structure of social networks
- Milgram observation – “no matter where we live, the world around us seems to be *small”*
- He showed that on average Americans are no more than six steps apart from each other (*six degrees of separation)*



Contd…

- *characteristic path length* - size of the average shortest path of the network *(called by milgram)*

- Geodesic (shortest path between two vertices v_s *and* v_t *) - a path that begins at the vertex* v_s *and ends in the vertex* v_t *and contains the least possible number of vertices*

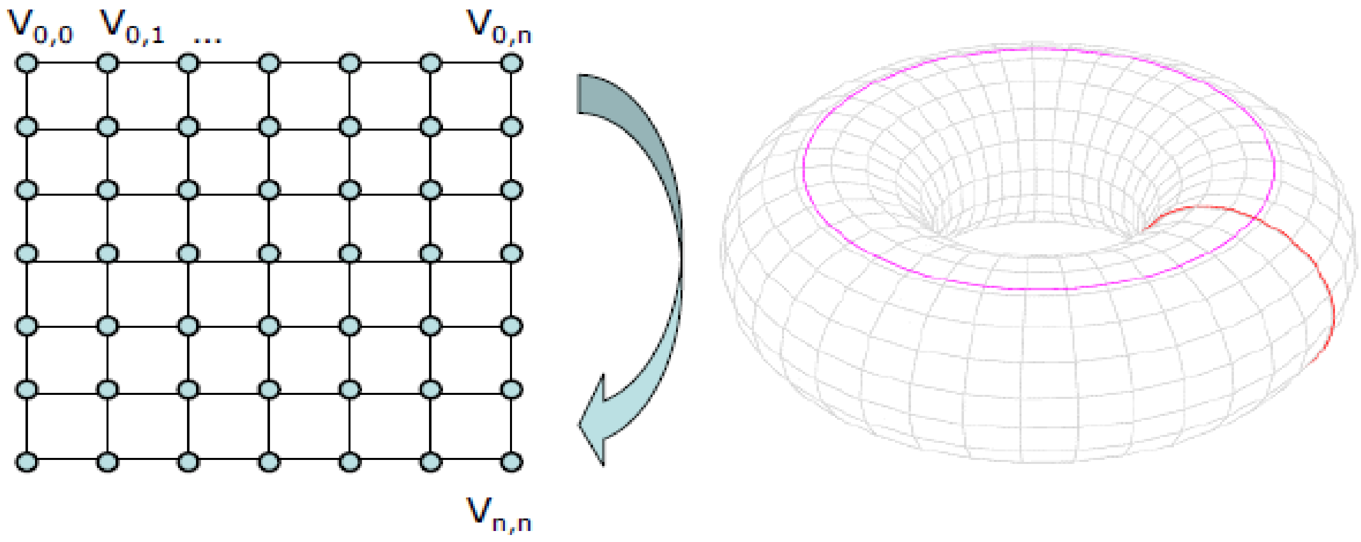
- Longest geodesic in the graph is called the diameter of the graph (maximum number of steps required between any 2 nodes)

- average shortest path - average of the length of the geodesics between all pairs of vertices in the graph

- Impossible for unconnected graph

Contd…

- Certain structures does not have small world property (2D lattice model) [characteristic path length of network size ‘n’ - 2/3 $*\sqrt{n}$]

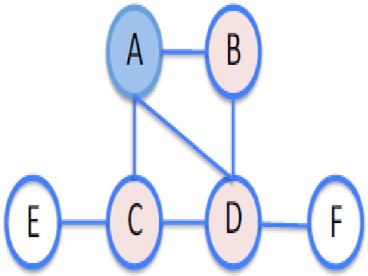


The 2D lattice model of networks (left). By connecting the nodes on the opposite borders of the lattice we get a toroidal lattice (right).

Clustering

- Friends are likely to know each other as well and tend to socialize in groups

- Other than that, other friends know each other because we introduced them to each other



- Clustering for a single vertex =

  Actual number of edges between neighbors of a vertex

  Possible number of possible edges between neighbors

  (no. of traiangles A is involved wrt ones it could be involved in)

- **averaging clustering over all vertices -> *clustering coefficient***
- *clustering coefficient = 0 (for a tree as there are no triads)*
- *It would never be the case that our friends are friends* with each other

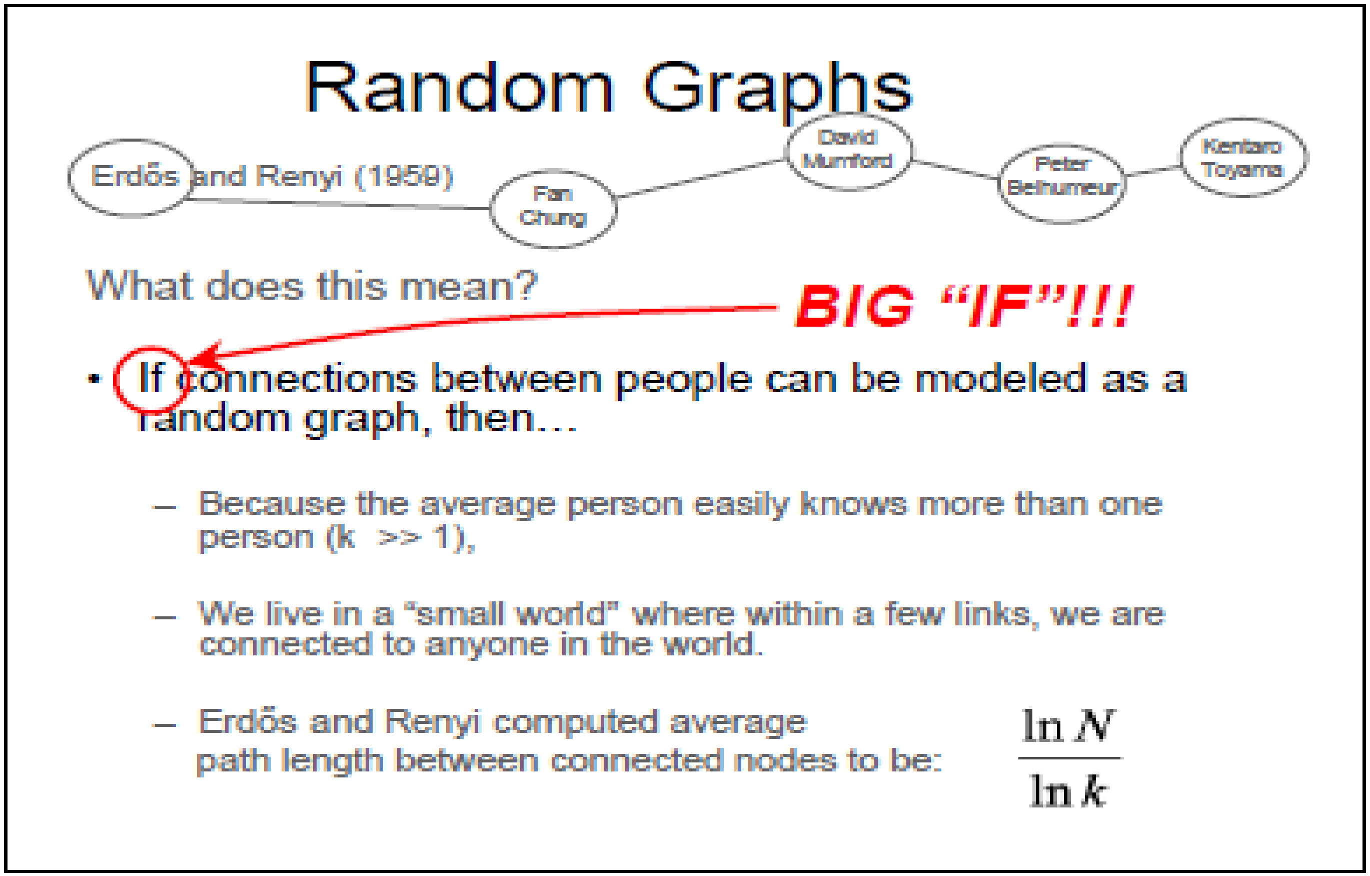
Random Graph Model

- Proposed by Hungarian mathematicians Erd˝os and R´enyi
- A random graph can be generated by taking a set of vertices with no edges connecting them
- Subsequently, edges are added by picking pairs of nodes with equal probability
- This way we create a graph where each pair of vertices will be connected with an equal probability
- higher probability—results in random graphs with small characteristic path length and most likely exhibit some **clustering**

Random Graphs

$p = 0.1$ $p = 0.25$ $p = 0.5$

Contd…


Random Graphs
Erdős and Renyi (1959)
Fan Chung
David Mumford
Peter Belhumeur
Kentaro Toyama
What does this mean?
BIG "IF"!!!
• If connections between people can be modeled as a random graph, then…
– Because the average person easily knows more than one person (k >> 1),
– We live in a "small world" where within a few links, we are connected to anyone in the world.
– Erdős and Renyi computed average path length between connected nodes to be:
$\frac{\ln N}{\ln k}$

Contd…

- ER random graphs are generative models – center point interest in knowing the process of growing complex networks in physics
- Researchers study the emergence of complex global structures from systems that are defined solely through elementary interactions

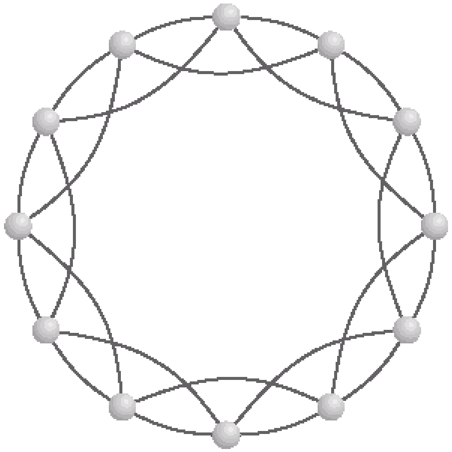
Alpha Model Networks

- In 1999, mathematicians Steven Strogatz and DuncanWatts presented alpha-model of networks
- Number of mutual friends shared by any two nodes determines the likelihood of a new tie
- Parameter alpha (number of friends) influence on the probability of a tie
- Alpha model generates graphs with small path lengths and relative large clustering coefficients

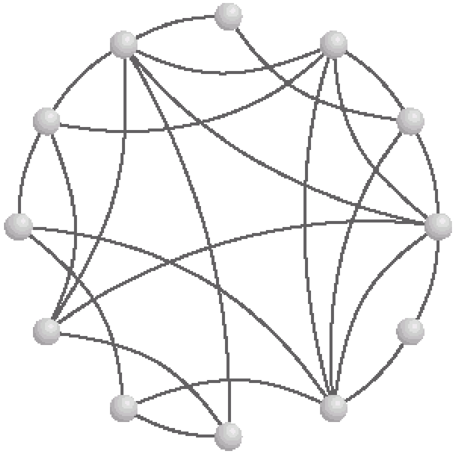
Beta Network Model

- alpha-model later simplified by authors so-called beta-model
- It achieves same results through less intuitive process
- It is Generative, one-dimensional toroidal lattice
- Every node connected to its neighbors and also its neighbors neighbors
- A random edge rewired keeping one end of edge fixed and other end rewired to another randomly selected node
- This process repeated until every link rewired with a probability of beta
- choosing beta appropriately, it can generate small path lengths and relatively large clustering coefficients network

Contd...



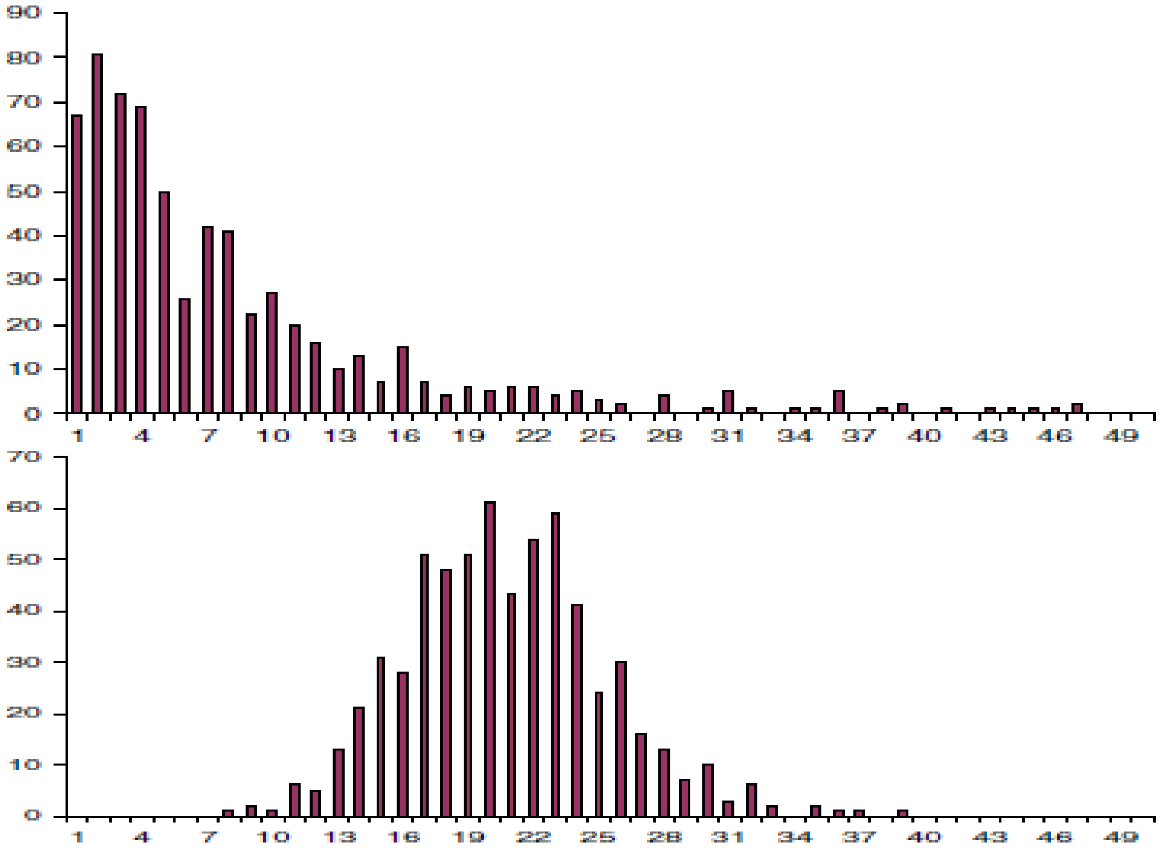
Rewired



Scale free Networks

- alpha and beta models fail to recreate scale-free characteristic of degree distribution
- Degree distribution of networks - diagram showing how many nodes in the network have a certain number of neighbors (degrees)
- In toroidal lattice, all nodes have an equal number of neighbors
- In alpha and beta models and random graphs shows normal distribution
- However, in real social networks, the higher the degree least likely to occur

Contd…



The degree distribution of a real world scale-free network (upper part) and the degree distribution of a random network of the same size (lower part)

Contd…

- steepness of the distribution shows - majority of nodes have much fewer connections except few *hubs of the network*
- *Barabasi discovered that exact correlation is a power law* - i.e. $p(d) = d^{-k}$ *, where* $k > 0$ *is a parameter of the distribution*
- He also gave a generative model to reproduce it

Contd…

- Barab´asi started with a single node and add nodes in every step
- when adding new nodes it is linked to the node to an already existing node
- probability of linking determined by how many edges the node already has (the rich get richer)
- a node that has already attracted more edges will have larger probability to attract connections in subsequent rounds
- Barab´asi showed that the resulting networks - have short characteristic path lengths, a large clustering coefficient and a degree distribution that approaches a power law