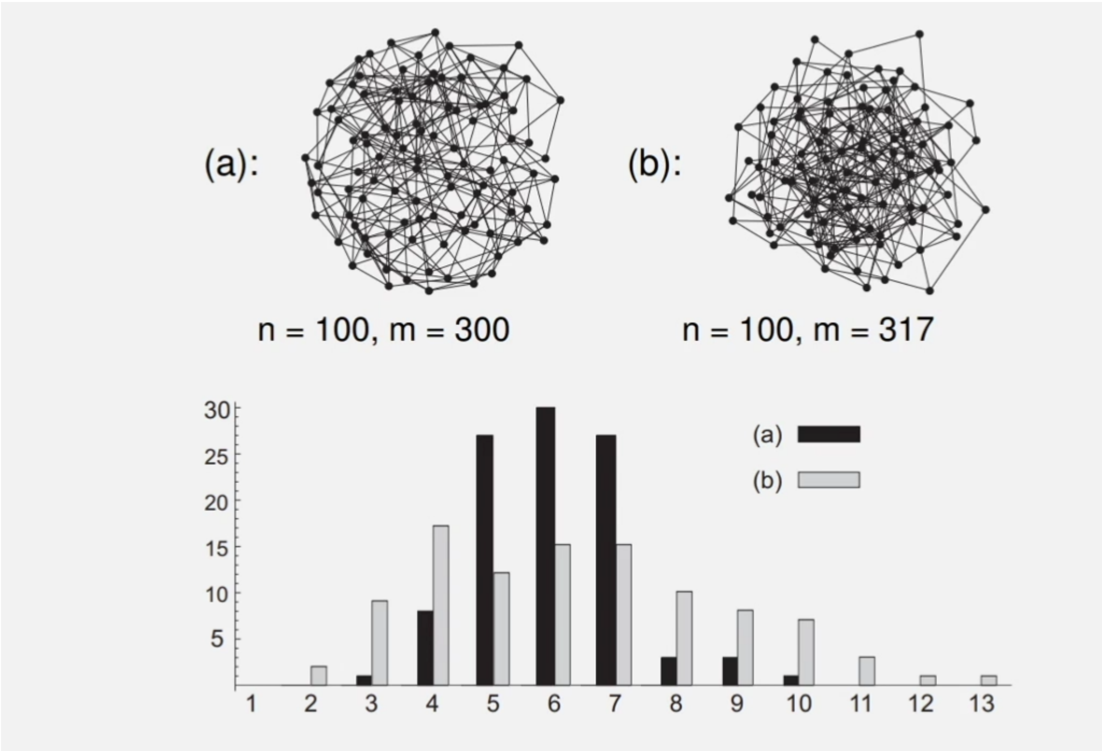
Networks and Graphs lecture 9

Network analysis

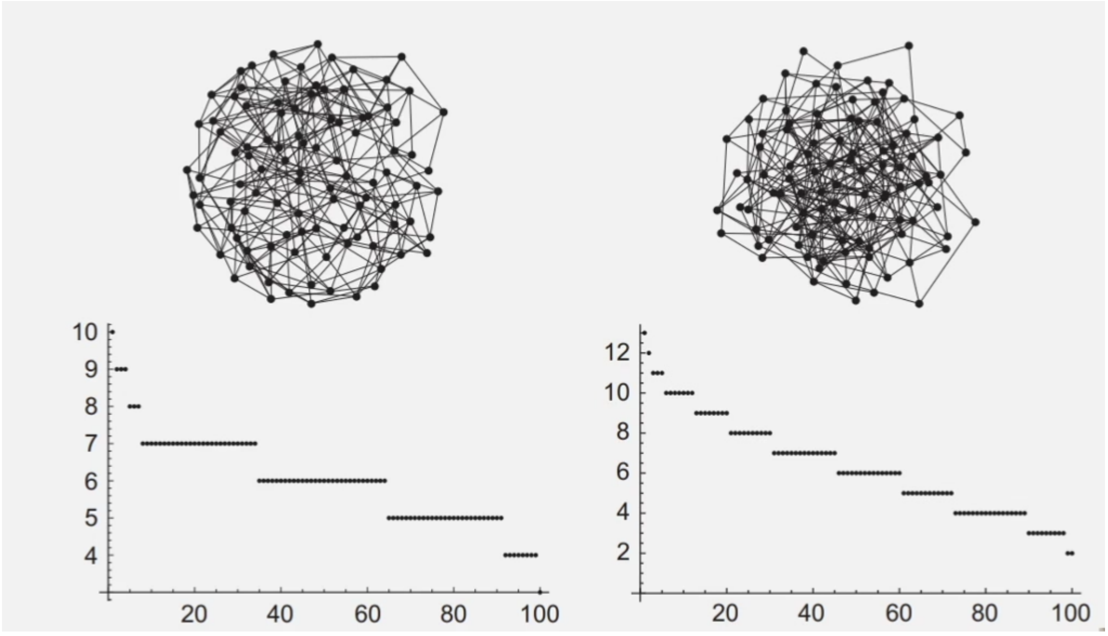
Distribution of vertex degrees

by creating a histogram from the nodes having some vertex degree, this way you can see more information at the same time.

> *vertex degree histogram:*
>
> 
>

by creating a ranked histogram you can also get more information about the vertex degrees.

> *ranked histogram of vertex degrees:*



Distance statistics

in this statistics only the shortest path will be considered and we just say the graphs are connected.

> *eccentricity: the maximum value of all shortest paths of all vertices from one node.*
>
> *radius: compute the eccentricity for all nodes in a graph, the minimum of those eccentricities is the radius of a graph.*
>
> *diameter: the maximum shortest path from any to any other node.*
>
> Eccentricity $\varepsilon(u)$: $\max\{d(u,v) \mid v \in V(G)\}$
> Radius $rad(G)$: $\min\{\varepsilon(u) \mid u \in V(G)\}$
> Diameter $diam(G)$: $\max\{d(u,v) \mid u, v \in V(G)\}$

average shortest path length

the average length is from one particular node to all other nodes

$$\frac{1}{|V(G)|-1} \sum_{v \in V(G)-\{u\}} d(u,v)$$

after computing this for all nodes you can compute the average path length.

$\bar{d}(G)$ denotes the average path length:

$$\frac{1}{|V(G)|}\sum_{u\in V(G)} \bar{d}(u)$$

characteristic path length

the characteristic path length is the median of all average length of all nodes.

Clustering

in real networks clustering will show up quite often, clustering is the vertices are neighbors of neighbors.

The maximum number of edges between neighbors of v is

$$(\delta(v)-1)+(\delta(v)-2)+\ \ldots+2+1=\frac{1}{2}\cdot\delta(v)\cdot(\delta(v)-1)$$

Alternatively, the maximum number of edges between neighbors of v can be represented as a *binomial coefficient*: $\binom{\delta(v)}{2}$.

clustering coefficient for one node:

m_v is the number of edges in the subgraph induced by $N(v)$: $|E(G[N(v)])|$

Definition

The clustering coefficient $cc(v)$ is defined by:

$$cc(v) \stackrel{\text{def}}{=} \begin{cases} \frac{m_v}{\binom{\delta(v)}{2}} = \frac{2\cdot m_v}{\delta(v)\cdot(\delta(v)-1)} & \text{if } \delta(v)>1 \\ \text{undefined} & \text{otherwise} \end{cases}$$

clustering coefficient for the **whole graph**:

Let $V^* \stackrel{\text{def}}{=} \{v \in V(G) | \delta(v) > 1\}$. The clustering coefficient $CC(G)$ is:

$$\frac{1}{|V^*|} \sum_{v \in V^*} cc(v)$$

the minimum value for the **cc** is 0 in case of a tree and the maximum is 1 if all neighbours of that vertex are connected.

triangles and triples

a triangle is as it says and a triple is a subgraph with 2 edges and 3 vertices, a triangle consists of 3 triples. this can be used to calculate the network transitivity.

if a triangle is found then the neighbors of that triangles origin are connected, and if it consists of a triple they are not connected.

A graph G has $n_\Delta(G)$ distinct triangles and $n_\wedge(G)$ distinct triples.
The network transitivity $\tau(G)$ is

$$\frac{n_\Delta(G)}{n_\wedge(G)}$$

the best value that can be obtained is 1.

- $n_\Delta(G) = \frac{1}{3} \cdot \sum_{v \in V^*} n_\Delta(v)$
- $n_\wedge(v) = \binom{\delta(v)}{2} = \frac{1}{2} \cdot \delta(v) \cdot (\delta(v) - 1)$
- $cc(v) = \frac{n_\Delta(v)}{n_\wedge(v)}$

Centrality

is the measuring if some nodes are more important than others. when a node is close to the center of the graph the eccentricity based vertex centrality is **high**.

The center $C(G)$ is the set of vertices with minimal eccentricity:

$$C(G) \stackrel{\text{def}}{=} \{v \in V(G) |\ \varepsilon(v) = rad(G)\}$$

$c_E(u)$ denotes the (eccentricity based) vertex centrality of u:

$$c_E(u) \stackrel{\text{def}}{=} \frac{1}{\varepsilon(u)}$$

a different way is by looking at closeness, this way you will concider all of the nodes and not only the maximum of the shortest path. also with the closeness the node will be central when the value is **high**.

$c_C(u)$ denotes the closeness of u:

$$c_C(u) \stackrel{\text{def}}{=} \frac{1}{\sum_{v \epsilon V(G)} d(u,v)}$$

another way to look at centrality is to look at the betweenness, this is when a node is on the shortest path from 2 different nodes. if this is the case often then the **betweenness centrality** is high.

$S(x, y)$ is set of shortest paths between x and y.
$S(x, u, y) \subseteq S(x, y)$ contains the shortest paths that pass through u.
$c_B(u)$ denotes the betweenness centrality of u:

$$c_B(u) \stackrel{\text{def}}{=} \sum_{x \neq u \neq y} \frac{|S(x, u, y)|}{|S(x, y)|}$$

cutting one of these nodes would increase the shortest paths of many other nodes and thus increasing their communication time.