

Chapter 4

Network Analysis

- 4.1 Introduction to Network Analysis
- 4.2 Centrality
- 4.3 Random Networks
- 4.4 Small World Networks

Introduction to Network Analysis

We study the properties of particular nodes or groups of nodes in a network as well as the properties of a network as a whole.

- ▶ The choice of a measure highly depends on the specific application scenario.
- ▶ Not every measure is suitable for every type of network.
- ▶ Some measures can be normalized, some cannot.

Network density

Consider a simple, undirected graph G with n vertices and m edges.

- ▶ The **mean degree** c of a vertex is

$$c = \frac{1}{n} \sum_{i=1}^n \delta(i) = \frac{2m}{n}$$

(Because if there are m edges in total then there are $2m$ ends of edges, and the number of ends of edges is equal to the sum of the degrees of all vertices)

- ▶ The **network density** $\rho(G)$ of G is defined as the fraction of these edges that are actually present:

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1}$$

- ▶ Note: the maximum possible number of edges in a simple graph is $\binom{n}{2} = \frac{n(n-1)}{2}$.

Network density II

- ▶ The density lies strictly in the range $0 \leq \rho \leq 1$.
- ▶ If the density ρ tends to a constant as $n \rightarrow \infty$, the network is said to be **dense**.
- ▶ A network in which $\rho \rightarrow 0$ as $n \rightarrow \infty$ is said to be **sparse**.

Example: it seems unlikely that the number of a persons friends will double solely because the population of the world doubles. Friendship networks are therefore regarded as sparse.

Average Path Length

Let G be a connected graph with vertex set V , and let $\bar{d}(u)$ denote the average length of the shortest paths from vertex u to any other vertex v in G :

$$\bar{d}(u) := \frac{1}{|V|-1} \sum_{v \in V, v \neq u} d(u, v)$$

The **average path length** $\bar{d}(G)$ is defined as

$$\bar{d}(G) := \frac{1}{|V|} \sum_{u \in V} \bar{d}(u) = \frac{1}{|V|^2 - |V|} \sum_{u, v \in V, u \neq v} d(u, v)$$

The **characteristic path length** of G is defined as the median over all $\bar{d}(u)$.

Clustering Coefficient

Consider a simple connected, undirected graph G and vertex $v \in V(G)$ with neighbor set $N(v)$. Let $n_v = |N(v)|$ and m_v be the number of edges in the subgraph induced by $N(v)$, i.e., $m_v = |E(G[N(v)])|$. The **clustering coefficient $cc(v)$ for vertex v** with degree $\delta(v)$ is defined as

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

Consider a simple connected graph G . Let V^* denote the set of vertices $\{v \in V(G) | \delta(v) > 1\}$. The **clustering coefficient $CC(G)$ for G** is defined as

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

Degree Correlation

Let G be a simple graph with degree sequence $d = [d_1, \dots, d_n]$ and adjacency matrix \mathbf{A} .

Let $V(G) = \{v_1, \dots, v_n\}$ be such that $\delta(v_i) = d_i$.

- The **degree correlation** of G is defined as:

$$r_{deg}(G) := \frac{\sum_{i=1}^n \sum_{j=i+1}^n ((d_i - \bar{d})(d_j - \bar{d}) \cdot \mathbf{A}[i,j])}{\sum_{i=1}^n (d_i - \bar{d})^2}$$

where \bar{d} denotes the average vertex degree, i.e., $\frac{1}{n} \sum_{i=1}^n d_i$.

- An alternative notation for the degree correlation in which the adjacency matrix is not used at all:

$$r_{deg}(G) := \frac{\sum_{e_{i,j}} ((d_i - \bar{d})(d_j - \bar{d}))}{\sum_{i=1}^n (d_i - \bar{d})^2}$$

where $e_{i,j} \in E(G)$ if and only if (1) there is an edge joining vertices v_i and v_j , and (2) $i > j$.

A note on Degree Correlation

The concept of degree correlation is based on the **Pearson correlation** (also known as **Pearson's correlation coefficient**):

$$r(x, y) = \frac{\frac{1}{n} \sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where x and y are two stochastic variables, for which we have a series of observation pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. By dropping the $1/n$ we can further simplify to

$$r(x, y) = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

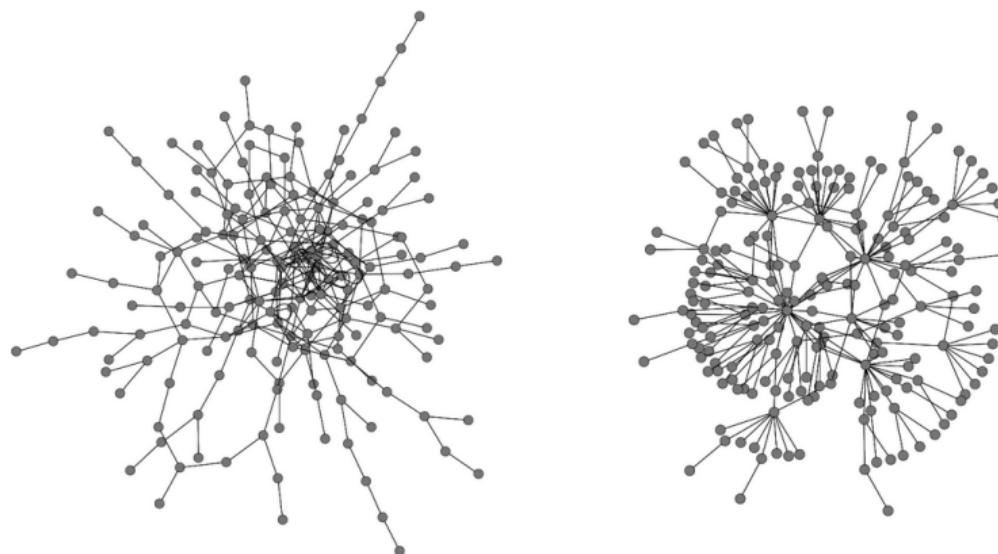
Assortative Mixing



Imagine a network in which every vertex is classified according to some characteristic that has a finite set of possible values. For example, the vertices could be people that are labelled according to their age, gender, or country of residence. Or they could represent drugs classified by undesirable side effects.

- ▶ The network is called **assortative** if a significant fraction of the edges in the network run between nodes of the same type.
- ▶ For instance, people have, it appears, a tendency to associate with others whom they perceive as being similar. Such a tendency is also called **homophily** or **assortative mixing**.
- ▶ Analogously, we define **disassortative mixing** as the tendency for nodes to associate with others who are *unlike* them.

Assortative mixing vs. Disassortative mixing



Two artificially generated networks.

(Left) Network that is assortative by degree. (Right) Disassortative network.¹

¹Source: Networks - An Introduction, M.E.J. Newman, Oxford University Press, 2010

Types of assortative mixing

In general, we can distinguish between two main types of assortative/disassortative mixing:

1. Assortative mixing by enumerative characteristics

- We consider a finite set of classes (labels) for the nodes. For example, genre, language, school class etc.

2. Assortative mixing by scalar characteristics

- We consider scalar characteristics such as age or income, i.e., values that are ordered. This way we can additionally say if two values are approximatively the same.
- A special case of assortative mixing by scalar characteristic is that of **mixing by degree**.

Assortative mixing by enumerative characteristics

Goal: the network is assortative if a significant fraction of the edges run between vertices of the same type.

- ▶ If we just measured that fraction, the result would not be very telling, because assortativity would be high also for trivial cases: for example, all friends of a human being are also human beings.
- ▶ Instead, we want a measure that is large for non-trivial cases and small for trivial ones.

Idea: we calculate the fraction of edges that run between vertices of the same type, and then we subtract from that figure the fraction of such edges we would *expect* to find if edges were positioned at random without regards to vertex type.

Assortative mixing by enumerative characteristics

- Let c_i be the class or type of vertex i , which is represented by an integer $1, \dots, n_c$, with n_c being the total number of classes.
- Then the total number of edges that are incident to vertices of the same type is given by

$$\sum_{\text{edges}(i,j)} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} \mathbf{A}_{ij} \cdot \delta(c_i, c_j)$$

where $\delta(x, y)$ is the Kronecker delta, which is defined as

$$\delta(x, y) := \begin{cases} 1 & \text{for } x = y \\ 0 & \text{for } x \neq y \end{cases}$$

and the factor of $1/2$ is required because every vertex pair is counted twice in the second sum.

Assortative mixing by enumerative characteristics

Next, we calculate the *expected* number of edges between vertices *if edges were placed at random*:

- ▶ Consider a particular edge attached to vertex i , which has degree k_i .
- ▶ By definition, we have m edges and $2m$ ends of edges in the network.
- ▶ The chance that our particular edge is one of the k_j ends attached to vertex j is $k_j/2m$ (if connections are made randomly).
- ▶ The total expected number of edges between vertices i and j is then $k_i k_j / 2m$.
- ▶ The **expected number of edges between all pairs of vertices of the same type** is given by

$$\frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j)$$

Modularity



Next, we consider the difference between the actual and expected number of edges in the network that join vertices of the same type:

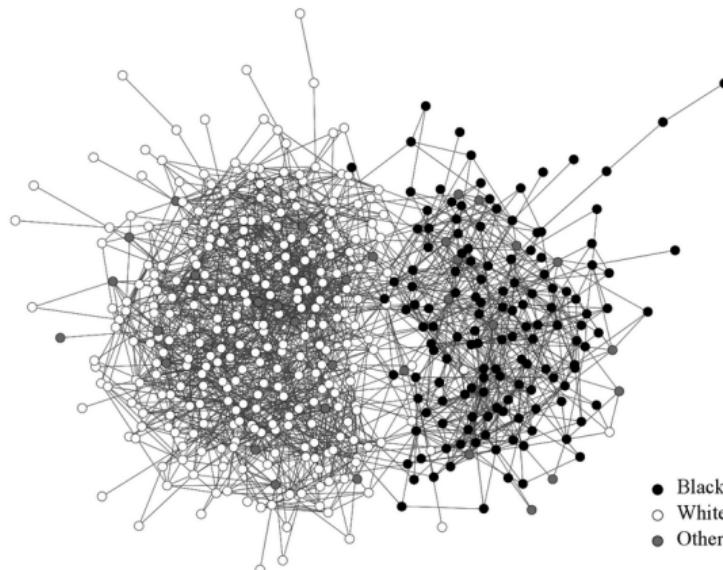
$$\frac{1}{2} \sum_{ij} \mathbf{A}_{ij} \cdot \delta(c_i, c_j) - \frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j) = \frac{1}{2} \sum_{ij} \left(\mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Generally, we are not so much interested in the absolute number of such edges, but rather the fraction, which is given by this same expression divided by the number m of edges:

$$Q = \frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

- ▶ This quantity Q is called the **modularity** and it is a measure of the extent to which like is connected to like in a network.
- ▶ Q is strictly less than 1, and it takes positive values if there are more edges between vertices of the same type than we would expect by chance, and negative ones if there are less.

Example: Assortative mixing by enumerative characteristics



Friendship network of 470 students (ages 14 to 18 years) at a US high school¹.
Modularity $Q = 0.305$

¹Source: Networks - An Introduction, M.E.J. Newman, Oxford University Press, 2010

Assortative mixing by scalar characteristics

Scalar characteristics of nodes are characteristics whose values come in a particular order. That way we can not only say that two characteristics are the same, but also say that they are approximately the same.

- ▶ For example: age, size, income, price

A common approach to determine assortative mixing by scalar characteristics is to use a covariance measure:

- ▶ Let x_i be the value for vertex i of the scalar value.
- ▶ Let (x_i, x_j) denote the pairs of values for the vertices at the ends of each edge $\langle i, j \rangle$.
- ▶ The mean μ of the value of x_i at the end of an edge is

$$\mu = \frac{\sum_{ij} \mathbf{A}_{ij} x_i}{\sum_{ij} \mathbf{A}_{ij}} = \frac{\sum_i k_i x_i}{\sum_i k_i} = \frac{1}{2m} \sum_i k_i x_i$$

Assortative mixing by scalar characteristics

Then, the covariance of x_i and x_j over edges is

$$\begin{aligned} \text{cov}(x_i, x_j) &= \frac{\sum_{ij} \mathbf{A}_{ij}(x_i - \mu)(x_j - \mu)}{\sum_{ij} \mathbf{A}_{ij}} \\ &= \frac{1}{2m} \sum_{ij} \mathbf{A}_{ij}(x_i x_j - \mu x_i - \mu x_j + \mu^2) \\ &= \left(\frac{1}{2m} \sum_{ij} \mathbf{A}_{ij} x_i x_j \right) - \mu^2 \\ &= \frac{1}{2m} \sum_{ij} \mathbf{A}_{ij} x_i x_j - \frac{1}{(2m)^2} \sum_{ij} k_i k_j x_i x_j \\ &= \frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j \end{aligned}$$

where we used the fact that $2m = \sum_{i=1}^n k_i = \sum_{ij} \mathbf{A}_{ij}$.

Assortative mixing by scalar value

- ▶ The covariance will be positive, if, overall, x_i and x_j at an incident edge are both large or both small and negative if they tend to vary in opposite directions.
- ▶ Often it is more convenient to normalize the covariance such that it is 1 in a perfectly mixed network (all edges fall between vertices with precisely equal values of x_i).
- ▶ If we put $x_j = x_i$ in the first term of the previous sum, we obtain a perfect mixing value of

$$\frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} x_i^2 - \frac{k_i k_j}{2m} x_i x_j \right) = \frac{1}{2m} \sum_{ij} \left(k_i \delta_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j$$

- ▶ The normalized measure, also called **assortativity coefficient**, is the ratio of the two:

$$r = \frac{\sum_{ij} (\mathbf{A}_{ij} - k_i k_j / 2m) x_i x_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) x_i x_j}$$

(This is in fact an example of a Pearson correlation coefficient.)

Assortative mixing by degree

Idea: high-degree vertices will be preferentially connected to other high-degree vertices, and the low to low.

- ▶ In such an assortative network we expect a clump or core of high-degree nodes surrounded by a less dense periphery.

We define a covariance similar as before, but with x_i as the degree k_i :

$$\text{cov}(k_i, k_j) = \frac{1}{2m} \sum_{ij} \left(\mathbf{A}_{ij} - \frac{k_i k_j}{2m} \right) k_i k_j$$

Again, we can normalize by the maximum value of the covariance to get an assortativity coefficient:

$$r = \frac{\sum_{ij} (\mathbf{A}_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j}$$

Centrality

One of the primary uses of graph theory in network analysis is the identification of the most important actors in a network.

- ▶ Important nodes are those that are extensively involved in relationships with other nodes, and thus, we define a **central** node as one involved in many ties:
 - ▶ If a node *receives* many ties, it is often said to be **prominent** or to have high **prestige**.
 - ▶ An actor who displays high *out-degree* centrality is often said to be **influential**.
- ▶ Centrality can also be applied at a group-level: These group-level-centrality measures allow us to compare different *networks* easily.

Degree Centrality

Perhaps the simplest centrality measure in a network is just the degree of a vertex, also called **degree centrality**.

$$k_i = \sum_{j=1}^n \mathbf{A}_{ij}$$

In directed networks, the **in-degree** is the number of ingoing edges connected to a vertex. The **out-degree** is the number of outgoing edges.

$$k_i^{in} = \sum_{j=1}^n \mathbf{A}_{ij}$$

$$k_j^{out} = \sum_{i=1}^n \mathbf{A}_{ij}$$

Eigenvector Centrality

Idea: extend the concept of degree centrality by increasing a vertex's importance if it is connected to other vertices that are *themselves important*.

- ▶ Goal: give each node a score proportional to the sum of the scores of its neighbors.
- ▶ We start by some initial guess about the centrality x_i of every vertex i . For example, we could set $x_i = 1$ for all i . Obviously this is not a useful measure, but we can use it to calculate a better one x'_i :

$$x'_i = \sum_j \mathbf{A}_{ij} x_j$$

- ▶ We can also write this expression as $x' = \mathbf{A}x$, where x is a vector with elements x_i .
- ▶ After t steps we have a vector of centralities $x(t)$ given by

$$x(t) = \mathbf{A}^t x(0)$$

Eigenvector Centrality

Let us write $x(0)$ as a linear combination of the eigenvectors v_i of the adjacency matrix:

$$x(0) = \sum_i c_i v_i$$

for some appropriate choice of constants c_i . Then

$$x(t) = \mathbf{A}^t \sum_i c_i v_i = \sum_i c_i \kappa_i^t v_i = \kappa_1^t \sum_i c_i \left[\frac{\kappa_i}{\kappa_1} \right]^t v_i$$

where the κ_i are the eigenvalues of \mathbf{A} , and κ_1 is the largest of them.

- ▶ Since $\kappa_i/\kappa_1 < 1$ for all $i \neq 1$, all terms in the sum other than the first decay as t becomes large, and hence in the limit $t \rightarrow \infty$ we get $x(t) \rightarrow c_1 \kappa_1^t v_1$.
- ▶ In other words, the limiting vector of centralities is simply proportional to the leading eigenvector of the adjacency matrix.
- ▶ Equivalently we could say that the centrality x satisfies

$$\mathbf{A}x = \kappa_1 x$$

This is the **eigenvector centrality**, first proposed by Bonacich in 1987.

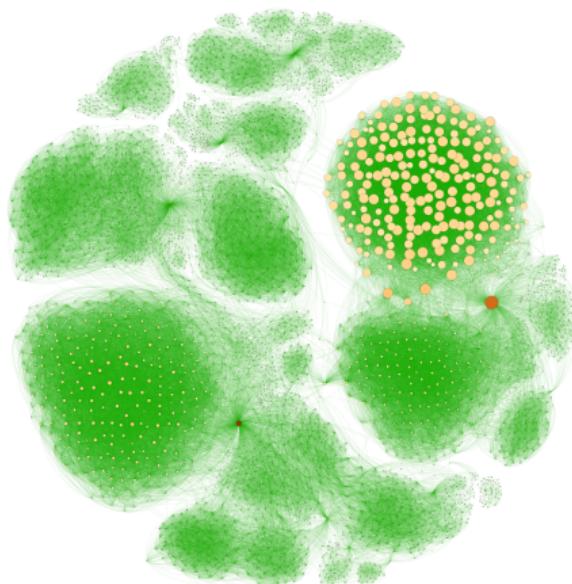
Eigenvector Centrality

- ▶ Thus the centrality x_i of vertex i is proportional to the sum of the centralities of i 's neighbors:

$$x_i = \kappa_1^{-1} \sum_j \mathbf{A}_{ij} x_j$$

- ▶ The eigenvector centrality has the nice property that it can be large either because a vertex has many neighbors or because it has important neighbors (or both).
- ▶ All eigenvector centralities of all vertices are non-negative.
- ▶ Eigenvector centrality works best for undirected networks.
- ▶ In directed networks some complications arise:
 - ▶ In general the adjacency matrix is asymmetric, this results in two leading eigenvectors. In most cases it is better to use the right eigenvector, given that centrality in directed networks is bestowed by other vertices pointing towards you.
 - ▶ A vertex with no incoming edges will always have a centrality of zero. This also affects its connected vertices, their centrality will also be zero. In mathematical term, only vertices in a *strongly connected component* will have non-zero eigenvector centrality.

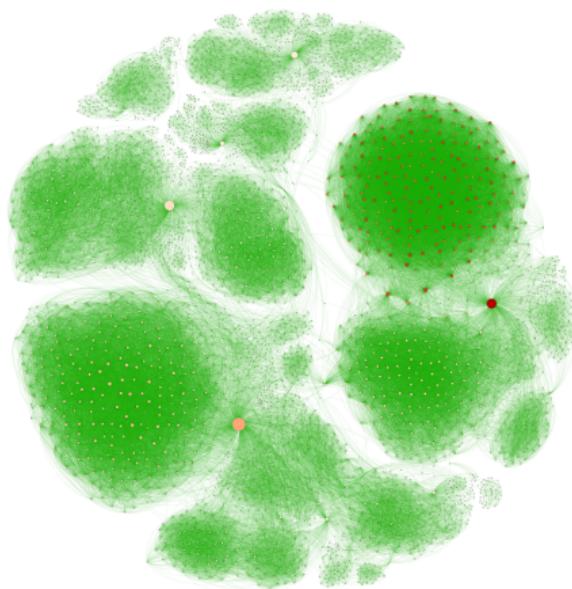
Eigenvector Centrality: Example 1



- ▶ Social network with $n = 4039$ nodes and $m = 88234$ edges¹.
- ▶ Node size indicates Eigenvector centrality.
- ▶ Node color indicates degree (more red = higher degree)

¹Data source: <https://snap.stanford.edu/>, image by Peter Ruppel

Eigenvector Centrality: example II



- ▶ Same network as before, however:
- ▶ Node size indicates degree.
- ▶ Node color indicates Eigenvector centrality (more red = higher Eigenvector centrality)

¹Data source: <https://snap.stanford.edu/>, image by Peter Ruppel

Eigenvectors and Eigenvalues

Let \mathbf{A} be a square matrix. A non-zero vector \mathbf{x} is called an **eigenvector** of \mathbf{A} if and only if there exists a number (real or complex) λ such that

$$\mathbf{Ax} = \lambda\mathbf{x}$$

If such a number λ exists, it is called an **eigenvalue** of \mathbf{A} . The vector \mathbf{x} is called eigenvector associated to the eigenvalue λ .

Katz Centrality

- With Eigenvector centrality, a vertex with no incoming edges will always have centrality zero, which also affects neighboring vertices. Example: a popular website with many incoming edges from websites that have centrality zero.
- Solution: give each vertex a small amount β of centrality for free:

$$x_i = \alpha \sum_j \mathbf{A}_{ij} x_j + \beta$$

Or in matrix notation:

$$\mathbf{x} = \alpha \mathbf{Ax} + \beta \mathbf{1}$$

where $\mathbf{1}$ is the vector $(1, 1, \dots, 1)^T$ and α and β are positive constants. Rearranging for \mathbf{x} , we find that

$$-\alpha \mathbf{Ax} + \mathbf{x} = \beta \mathbf{1}$$

$$(\mathbf{I} - \alpha \mathbf{A})\mathbf{x} = \beta \mathbf{1}$$

$$(\mathbf{I} - \alpha \mathbf{A})^{-1} (\mathbf{I} - \alpha \mathbf{A})\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A})^{-1} \beta \mathbf{1}$$

$$\mathbf{x} = (\mathbf{I} - \alpha \mathbf{A})^{-1} \beta \mathbf{1}$$

Katz Centrality

Given that normally we don't care about the absolute magnitude, we usually set $\beta = 1$, giving the **Katz centrality**:

$$x = (\mathbf{I} - \alpha \mathbf{A})^{-1} \mathbf{1}$$

Observations:

- ▶ Which value to choose for α ? For $\alpha \rightarrow 0$, all vertices have the same centrality β (which we have set to 1). As we increase α from zero the centrality increases and eventually there comes a point where they diverge. This happens where $(\mathbf{I} - \alpha \mathbf{A})^{-1}$ diverges, i.e., when $\det(\mathbf{I} - \alpha \mathbf{A})$ passes through zero:

$$\det(\mathbf{A} - \alpha^{-1} \mathbf{I}) = 0$$

As α increases, the determinant first crosses zero when α^{-1} equals the largest eigenvalue of \mathbf{A} . Therefore we should choose a value of α less than $1/\kappa_1$ because then we can be sure that the corresponding matrix is invertible.

- ▶ Most researchers employ values of α close to $1/\kappa_1$.

Katz Centrality: Observations

- ▶ Calculating the exact Katz centrality comes down to finding the inverse of $(\mathbf{I} - \alpha \mathbf{A})$. Inverting a matrix takes $\mathcal{O}(n^3)$ time, which is not feasible for large matrices.
- ▶ There exist several approximation techniques for inverting a matrix \mathbf{M} , for example, we can use a Neumann series

$$\mathbf{M}^{-1} = \sum_{k=0}^{\infty} (\mathbf{I} - \mathbf{M})^k$$

and stop after an appropriate number of iterations.

- ▶ However, in practice, the Katz centrality is usually calculated analogously to the Eigenvector centrality, i.e., by iteratively calculating the individual centralities until the change between rounds becomes small enough. In other words, we make an initial estimate of x (probably a bad one such as $x = 0$) and calculate a better one by

$$x' = \alpha \mathbf{Ax} + \beta \mathbf{1}$$

PageRank

Drawback of the Katz centrality:

- ▶ A high-centrality vertex pointing to many others gives the others also high centrality.
Example: a search engine might contain a link to my webpage, but it also has links to millions of other pages. Does this make my webpage central or not?

To overcome this issue we introduce the following concept:

$$x_i = \alpha \sum_j \mathbf{A} \frac{x_j}{k_j^{out}} + \beta$$

- ▶ Simply put: *The centrality I derive from my neighbors is proportional to their centrality divided by their out-degree.*
- ▶ This is also known as the **PageRank**.
- ▶ We require k_j^{out} to be set to 1 in case the vertex has no outgoing edges.

PageRank

We can write

$$x_i = \alpha \sum_j \mathbf{A} \frac{x_j}{k_j^{out}} + \beta$$

in matrix notation:

$$x = \alpha \mathbf{AD}^{-1}x + \beta \mathbf{1}$$

with \mathbf{D} being the diagonal matrix with elements $D_{ii} = \max(k_i^{out}, 1)$. Usually we don't care about absolute values, so we set $\beta = 1$ and find that

$$\begin{aligned} x &= (\mathbf{I} - \alpha \mathbf{AD}^{-1})^{-1} \mathbf{1} \\ &= (\mathbf{DD}^{-1} - \alpha \mathbf{AD}^{-1})^{-1} \mathbf{1} \\ &= [(\mathbf{D} - \alpha \mathbf{A}) \mathbf{D}^{-1}]^{-1} \mathbf{1} \\ &= \mathbf{D}(\mathbf{D} - \alpha \mathbf{A})^{-1} \mathbf{1} \end{aligned}$$

(Whereby in the last step we used that $(\mathbf{MN})^{-1} = \mathbf{N}^{-1} \mathbf{M}^{-1}$ for two (invertible) matrices \mathbf{M} and \mathbf{N})

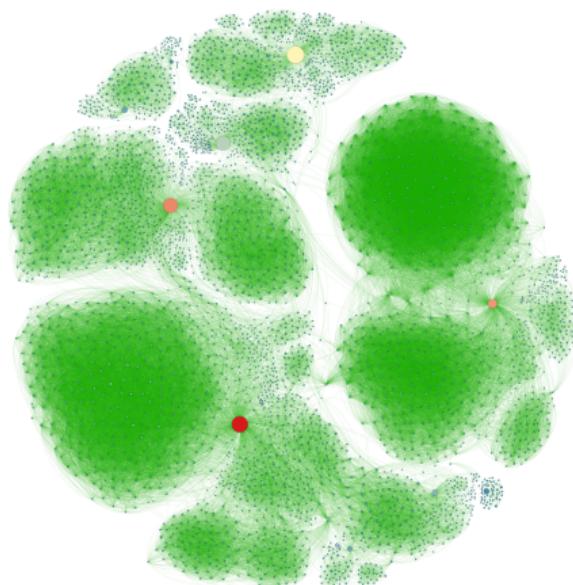
PageRank

- ▶ For PageRank, α should be less than the inverse of the largest eigenvalue of \mathbf{AD}^{-1} .
For an undirected network, this largest eigenvalue turns out to be 1 with corresponding eigenvector k_1, k_2, \dots, k_n .
- ▶ For a directed network in practical cases the leading eigenvalue will be roughly of order 1.
- ▶ For example, a common value is $\alpha = 0.85$.

In the original paper, Brin and Page defined the PageRank as follows: We assume page A has pages T_1, \dots, T_n which point to it (i.e., are citations). The parameter d is a dampening factor which can be set between 0 and 1. We usually set d to 0.85. [...] Also $C(A)$ is defined as the number of links going out of page A . The PageRank of page A is given as follows:

$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right)$$

PageRank: Example



- ▶ Social network with $n = 4039$ nodes and $m = 88234$ edges¹.
- ▶ Node size indicates PageRank.
- ▶ Node color indicates degree (more red = higher degree)

¹Data source: <https://snap.stanford.edu/>, image by Peter Ruppel

Measures based on shortest paths

Another way of looking for central nodes is to consider distances, i.e., shortest paths between nodes:

- ▶ The **closeness centrality** measures the mean distance from a vertex to other vertices.
- ▶ The **betweenness centrality** measures the extent to which a vertex lies on paths between other vertices.

Closeness Centrality

- ▶ Let G be a directed or undirected graph and $u, v \in V(G)$ with $|V(G)| \geq 2$.
- ▶ Let $d(u, v)$ be the length of a shortest (u, v) -path, also called **(geodesic) distance** between u and v . Note that the shortest path itself does not need to be unique, i.e., there can be different shortest paths.
- ▶ The **mean geodesic distance** \bar{d}_i between a particular vertex i and all other $n - 1$ vertices is given by

$$\bar{d}_i = \frac{1}{n-1} \sum_j d(i, j)$$

This measure takes low values for vertices that are "better connected".

- ▶ A more common approach is to use the inverse, also called the **closeness centrality**:

$$C_C(i) = \frac{n-1}{\sum_j d(i, j)}$$

Betweenness Centrality

The **betweenness centrality** measures the extent to which a vertex lies on paths between other vertices. A high betweenness indicates, e.g., that

- ▶ the vertex may have considerable influence within a network
- ▶ many messages could pass through the vertex
- ▶ the removal of that vertex will most probably disrupt communications between other vertices

Let σ_{st} denote the number of shortest paths from $s \in V$ to $t \in V$. Let $\sigma_{st}(v)$ denote the number of shortest paths from s to t that some $v \in V$ lies on. Then the **betweenness centrality** is given by

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Betweenness Centrality

Usually, we normalize to obtain values between 0 and 1. The highest possible value occurs when one node is located on every single shortest path ("star graph").

- ▶ For *undirected* networks the **normalized betweenness centrality** is given by

$$C_B(v) = \frac{2}{n^2 - 3n + 2} \cdot \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

i.e, we divided the betweenness centrality by $(n - 1)(n - 2)/2$.

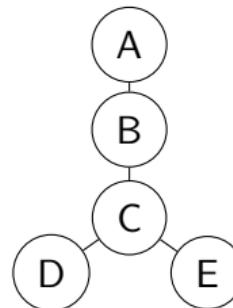
- ▶ For *directed* networks the **normalized betweenness centrality** is given by

$$C_B(v) = \frac{1}{n^2 - 3n + 2} \cdot \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

i.e, we divided the betweenness centrality by $(n - 1)(n - 2)$.

Betweenness Centrality: example

Consider the following undirected network:



	$C_B(v)$	Normalized $C_B(v)$
A	0	0
B	3.0	0.5
C	5.0	0.83
D	0	0
E	0	0

How to calculate the betweenness centrality?

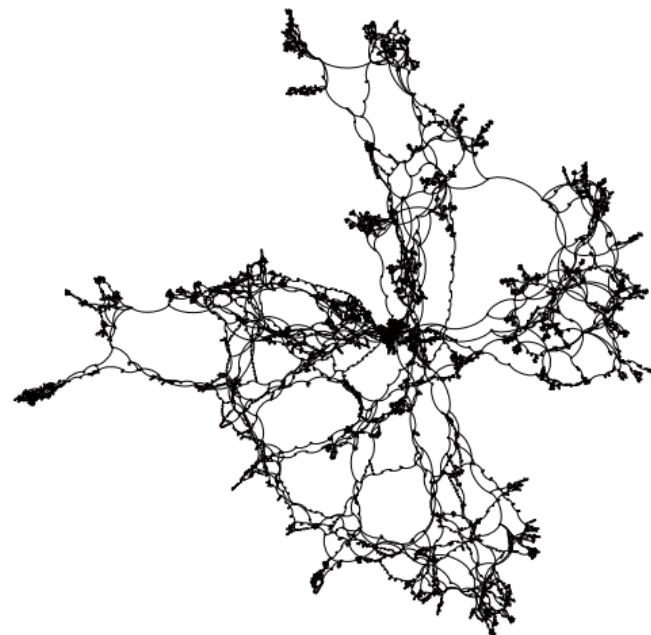
- ▶ **Brandes' algorithm**¹ finds the betweenness centrality in $\mathbf{O}(nm \times n^2 \log(n))$ runtime for weighted graphs and requires $\mathbf{O}(n + m)$ space. For unweighted graphs, runtime reduces to $\mathbf{O}(nm)$.
- ▶ **Nasre's algorithm**² runs in $\mathbf{O}(m'n + n^2)$ time, where m' is the size of a certain subset of E , the set of edges in G that lie on a shortest path.

¹Ulrik Brandes, A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology 25(2):163-177, 2001

²Meghana Nasre, Matteo Pontecorvi, and Vijaya Ramachandran, Betweenness Centrality – Incremental and Faster, Mathematical Foundations of Computer Science 2014, Volume 8635 of the series Lecture Notes in Computer Science pp 577-588

Comparison of measures

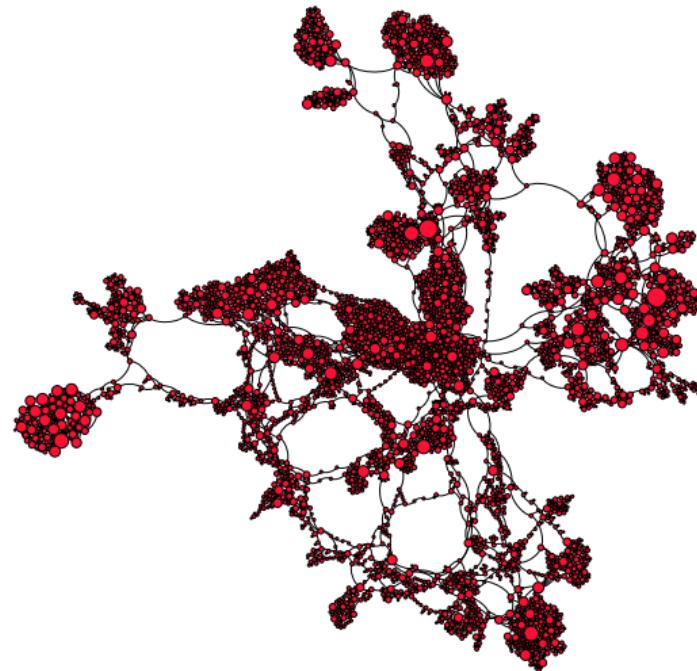
Consider the following undirected network, which represents a power grid¹ and consists of 4941 nodes and 6594 edges:



¹Data source: D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998) / <https://github.com/gephi/gephi>

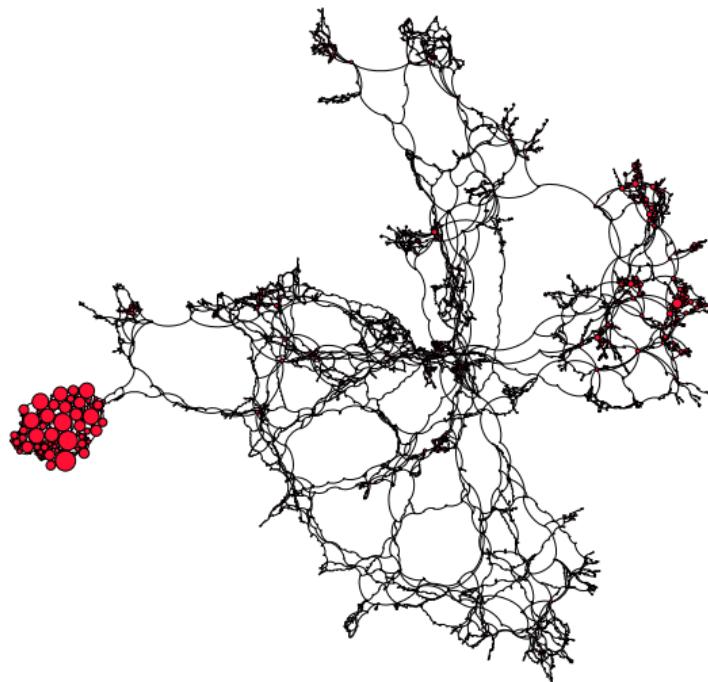
Comparison of measures

Node size represents the degree:



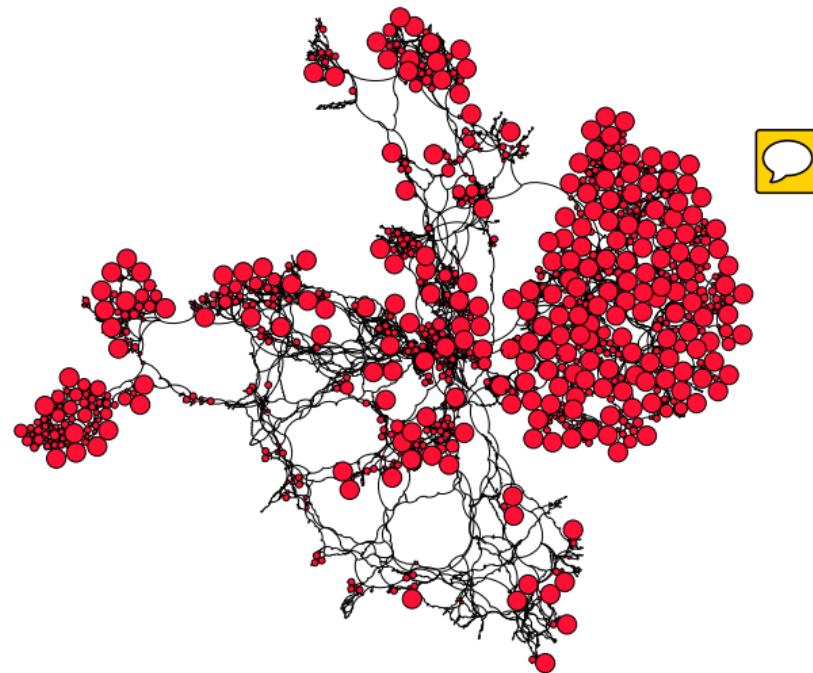
Comparison of measures

Node size represents Eigenvector centrality:



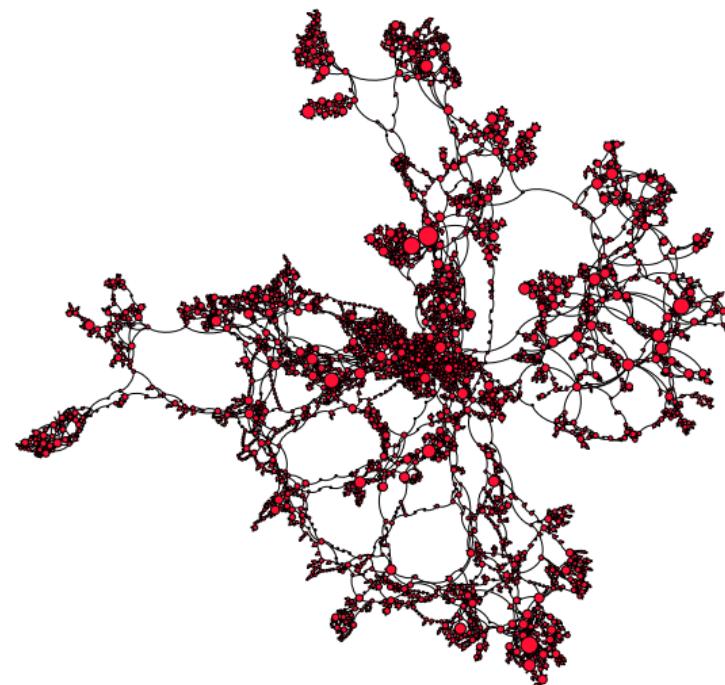
Comparison of measures

Node size represents clustering coefficient:



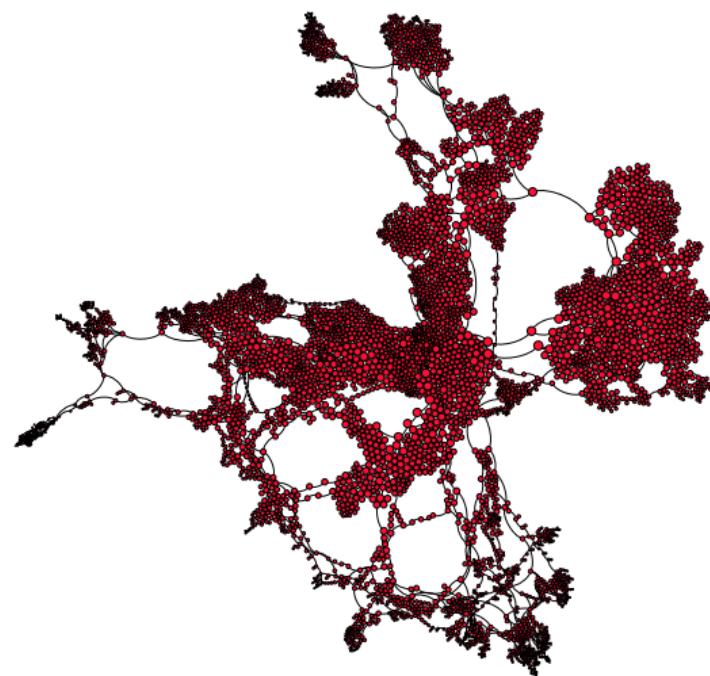
Comparison of measures

Node size represents PageRank:



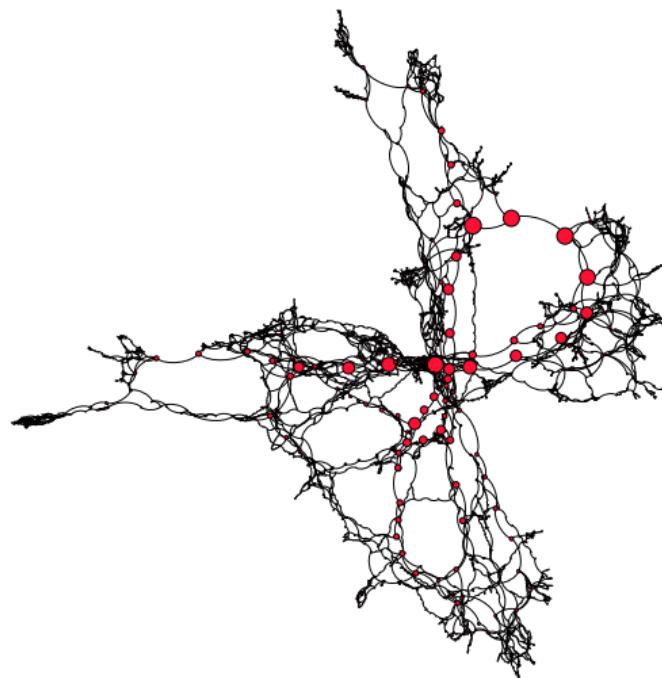
Comparison of measures

Node size represents closeness centrality:



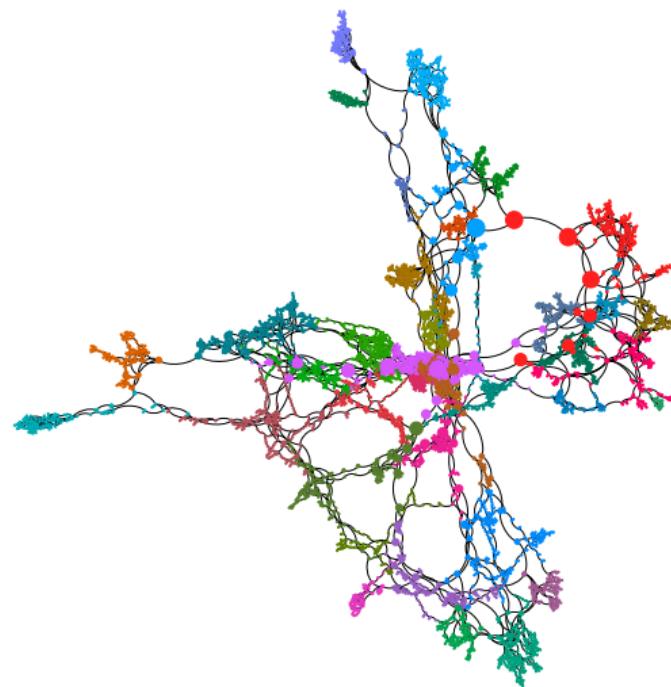
Comparison of measures

Node size represents betweenness centrality:



Comparison of measures

Node size represents betweenness centrality (the minimum size has been increased slightly to improve visibility), node color indicates modularity class:



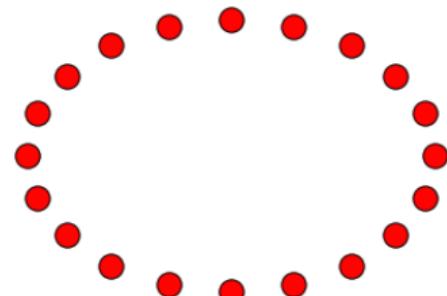
Random Networks

- ▶ So far we have considered the structure of networks as well as mathematical methods for making sense of the network data.
- ▶ Another way of looking at networks is to consider some fixed properties such as a degree distribution and to ask what wider effect it will have on the behavior of the system.
- ▶ In general, a **random graph** is a model network in which some specific set of parameters take fixed values, but the network is random in other respects.
- ▶ For example, we take n vertices and place m edges among them at random (typically we require the graph to remain simple). Such a graph is typically referred to by $G(n, m)$. Equivalently, we could say, that G is created by choosing uniformly at random among the set of all simple graphs with exactly n vertices and m edges.

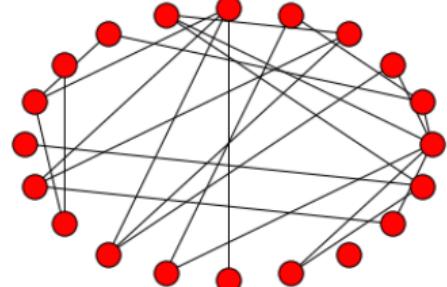
Erdös Rényi Model (Random Network)

- ▶ $G(n, p)$ is also commonly known as an Erdös Rényi model.
- ▶ An **Erdös Rényi model** of a random network on n vertices, also referred to as an **ER random graph**, is an undirected graph $G_{n,p}$ in which each two distinct vertices are connected by an edge with probability p . For a given number M of edges, the ER random graph $G_{n,M}$ is an undirected graph in which each of the M edges is incident to randomly chosen pairs of vertices.
- ▶ The set of *all* ER random graphs, with n vertices and probability p that two vertices are joined, is denoted by $ER(n, p)$.
- ▶ The clustering coefficient of any $ER(n, p)$ is equal to p .

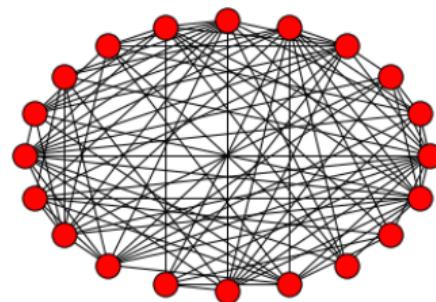
Examples for ER random graphs



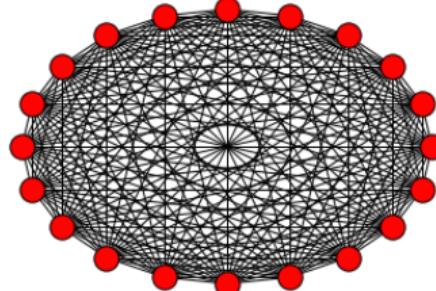
$n = 20, p = 0.0$



$n = 20, p = 0.1$



$n = 20, p = 0.5$



$n = 20, p = 1.0$

Watts-Strogatz Random Graph

Consider a set on n vertices $\{v_1, v_2, \dots, v_n\}$ and an (even) number k . In order to ensure that the graph will have relatively few edges (i.e, it is sparse), choose n and k such that $n \gg k \gg \ln(n) \gg 1$.

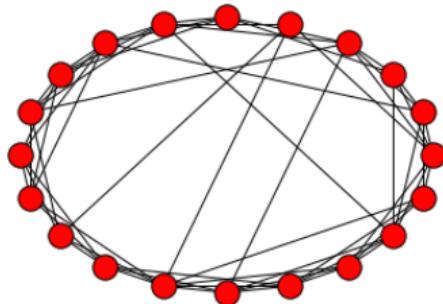
1. Order the n vertices into a ring and connect each vertex to its first $k/2$ left-hand (clockwise) neighbors, and its $k/2$ right-hand (counterclockwise) neighbors, leading to graph G .
2. With probability p , replace each edge $\langle u, v \rangle$ with an edge $\langle u, w \rangle$ where w is a randomly chosen vertex from $V(G)$ other than u , and such that $\langle u, w \rangle$ is not already contained in the edge set of (the modified) G .

The resulting graph is called a **Watts-Strogatz random graph** (or simply a **WS graph**). We also refer to it as a $WS(n, k, p)$ graph.

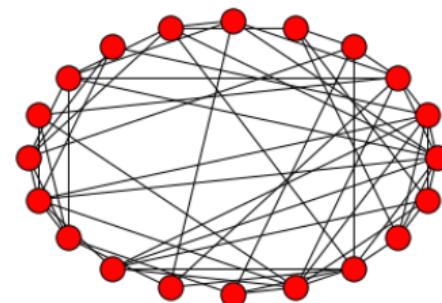
Examples for WS random graphs



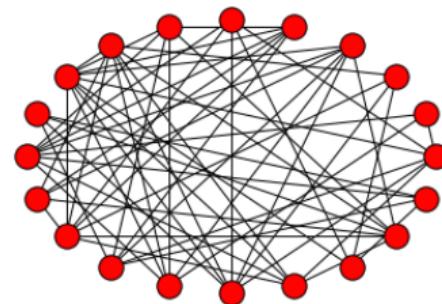
$n = 20, k = 6, p = 0.0$



$n = 20, k = 6, p = 0.1$



$n = 20, k = 6, p = 0.5$



$n = 20, k = 6, p = 1.0$

WS graphs

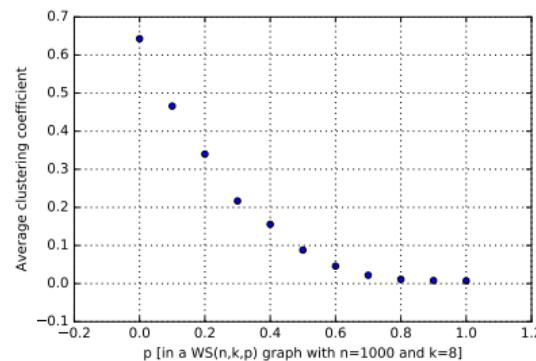
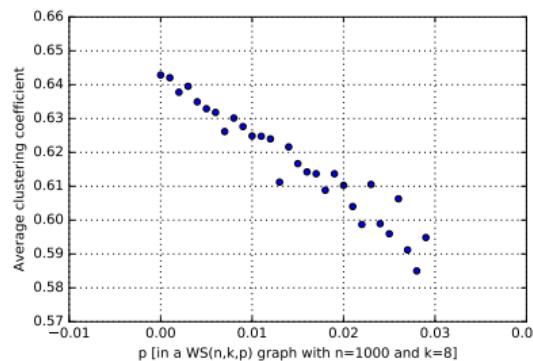
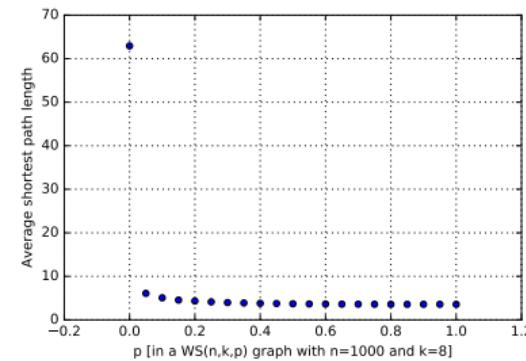
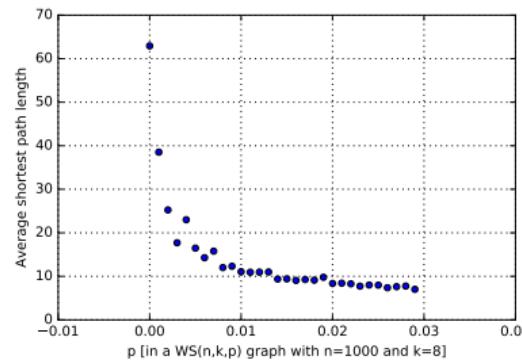
For a Watts-Strogatz graph G from $WS(n, k, 0)$ the average shortest path length $\bar{d}(u)$ from a given vertex u to any other vertex in G is approximated by

$$\bar{d}(u) \approx \frac{(n - 1)(n + k - 1)}{2kn}$$

In other words, $WS(n, k, 0)$ may show a high clustering coefficient, yet it does not have small average shortest path lengths.

- ▶ However, if we slightly increase p , the average path length drops rapidly.

WS graphs: average shortest path length vs. average clustering coefficient



Small World Networks

In 1967 Stanley Milgram published a study in which he tried to answer the question "Given any two people in the world, person X and person Z, how many intermediate acquaintance links are needed before X and Z are connected?" (where acquaintance refers to a person who knows both X and Z).

- ▶ Milgram sent letters to $n = 296$ random participants and asked them to forward their letter toward a target.
- ▶ First study: start in Wichita, Kansas and end at the wife of a divinity school student in Cambridge.
- ▶ Second study: start in Omaha, Nebraska and end at a stockbroker who worked in Boston and lived in Sharon, Massachusetts.

Small World Networks

Jeffrey Travers and Stanley Milgram, An Experimental Study of the Small World Problem, *Sociometry Vol. 32, No. 4 (Dec., 1969)*, pp. 425-443:

[...] *The procedure may be summarized as follows: an arbitrary "target person" and a group of "starting persons" were selected, and an attempt was made to generate an acquaintance chain from each starter to the target. Each starter was provided with a document and asked to begin moving it by mail toward the target. The document described the study, named the target, and asked the recipient to become a participant by sending the document on. It was stipulated that the document could be sent only to a first-name acquaintance of the sender. The sender was urged to choose the recipient in such a way as to advance the progress of the document toward the target; several items of information about the target were provided to guide each new sender in his choice of recipient. Thus, each document made its way along an acquaintance chain of indefinite length, a chain which would end only when it reached the target or when someone along the way declined to participate. [...]*

Results of Milgram's experiment

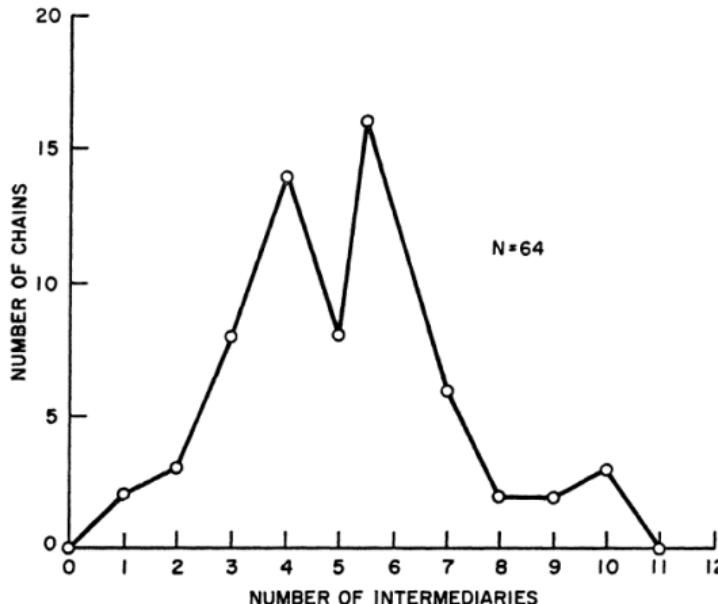


FIGURE 1
Lengths of Completed Chains

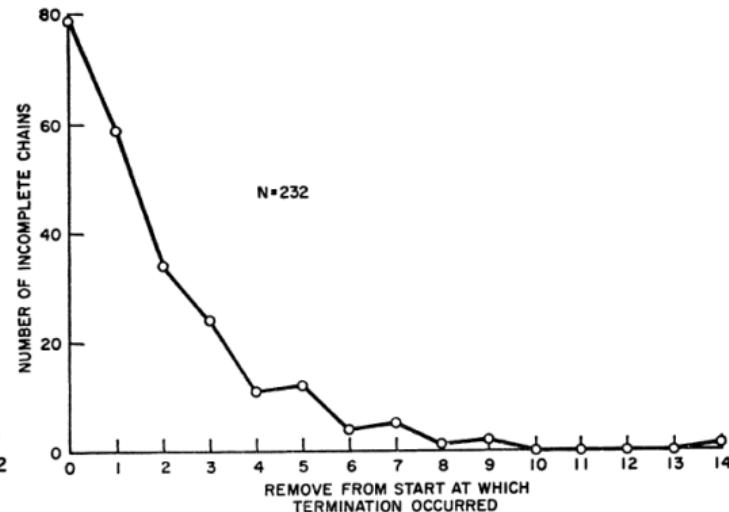


FIGURE 2
Lengths of Incomplete Chains

Source: Jeffrey Travers and Stanley Milgram, An Experimental Study of the Small World Problem, *Sociometry* Vol. 32, No. 4 (Dec., 1969), pp. 425-443

Current example: Facebook

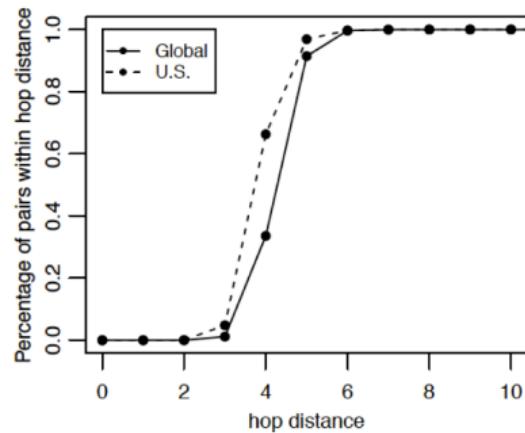


Figure 2. Diameter. The neighborhood function $N(h)$ showing the percentage of user pairs that are within h hops of each other. The average distance between users on Facebook in May 2011 was 4.7, while the average distance within the U.S. at the same time was 4.3.

Source: Johan Ugander, Brian Karrer, Lars Backstrom, Cameron Marlow: The Anatomy of the Facebook Social Graph, 2011

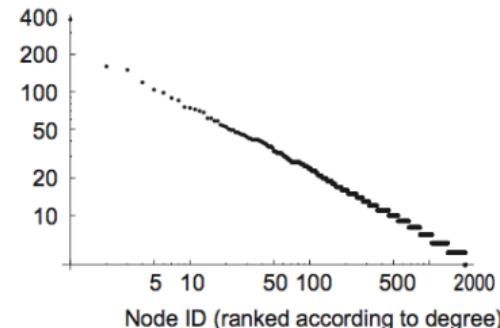
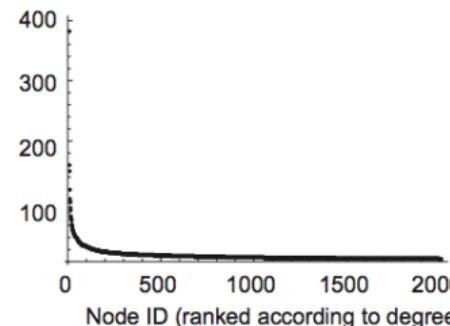
Small World Networks

- ▶ What Milgram demonstrated, and what has been shown to hold in many real-world situations, is that the average shortest path length is relatively small.
- ▶ Small world networks show properties of ER random graphs, however they differ in terms of their clustering coefficient and average shortest path length.
- ▶ Watts-Strogatz random graphs are generally considered to represent small-world phenomenon. However, WS random graphs often do not capture (other) properties of real-world networks, such as communication networks or biological networks.
- ▶ Albert-László Barabási and Réka Albert showed that real-world networks such as the World Wide Web, actor collaborations, and many more, exhibit a structure in which there are *a few high-degree nodes*, but that the number of nodes with a high degree *decreases exponentially*.

Scale-free Networks

A network is called **scale-free** if the distribution of vertex degrees follows a **power law**, which means that the probability that an arbitrary node has degree k is proportional to $(1/k)^\alpha$ for some number $\alpha > 1$ (called the **scaling exponent**).

- ▶ Example: the distribution of vertex degrees of a scale-free network with 2000 nodes, shown as (a) a linear plot and as (b) a log-log plot:



- ▶ Typical values for α are in the range $2 \leq \alpha \leq 3$.

Scale-free Networks

- ▶ Scale-free graphs are fundamentally different to ER and WS random graphs, because it appears that we can construct them only through a growth process combined with what is referred to as preferential attachment.
- ▶ Barabási and Albert were the first to devise a procedure for the construction of scale-free networks.

Scale-free Networks

A scale-free network can be artificially created, e.g., by the algorithm of Barabási and Albert: Consider a (relatively small) ER random graph G_0 with n_0 vertices V_0 . At each step $s > 0$:

1. Add a new vertex v_s to V_{s-1} (i.e., $V_s \leftarrow V_{s-1} \cup \{v_s\}$).
2. Add $m \leq n_0$ edges to the graph, each edge being incident with v_s and a vertex u from V_{s-1} chosen with probability

$$P[\text{select } u] = \frac{\delta(u)}{\sum_{w \in V_{s-1}} \delta(w)}$$

that is, choosing a vertex u is proportional to the current vertex degree of u . Vertex u must not have been previously chosen during this step.

3. Stop when n vertices have been added, otherwise repeat the previous two steps.

The resulting graph is called a **Barabási-Albert random graph**, also denoted by $BA(n, n_0, m)$.