Graphs 1: Introduction to graphs and their parameters

300958 Social Web Analytics

WESTERN SYDNEY
UNIVERSITY

School of Computing, Engineering and Mathematics

Week 7

Outline

Social Networks

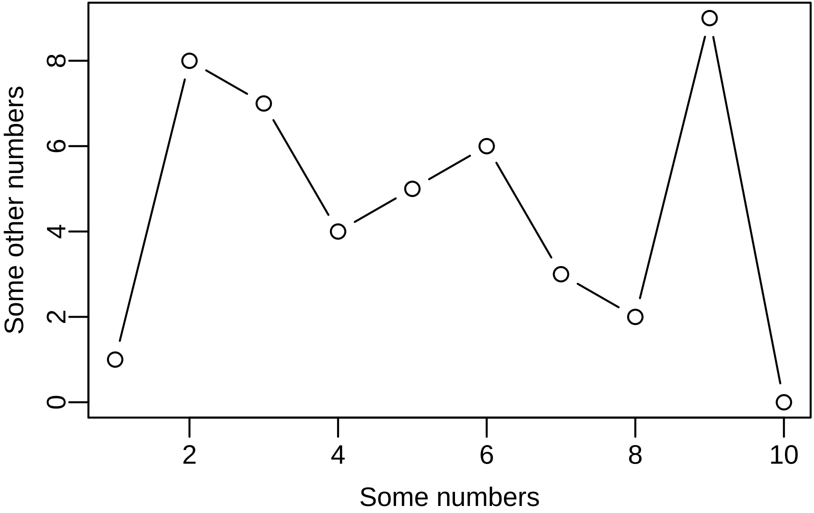
It is the *Network* that makes Social Networking sociable. Without the social network, the Web would be less dynamic. There would be less information flow, meaning that the Web would be less useful.

A network is also called a graph. In this lecture we will examine basic graph theory, to allow us to analyse the social network.

Outline

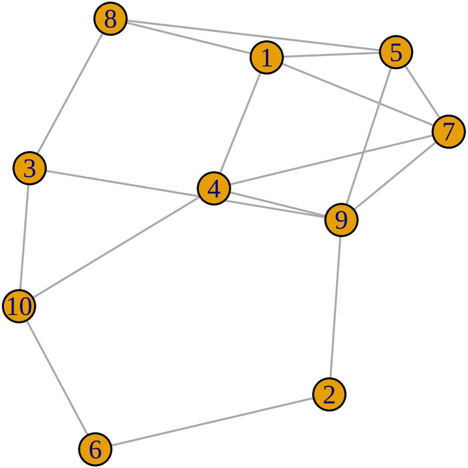
Not a graph

This was a graph when you were in high school. This is now called a *plot*.



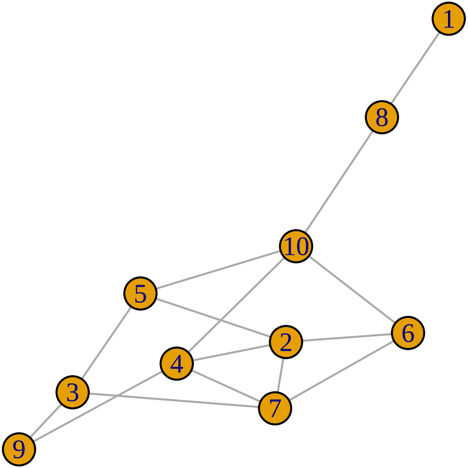
A graph

Now that you have passed high school, we can let you know that this is a graph.



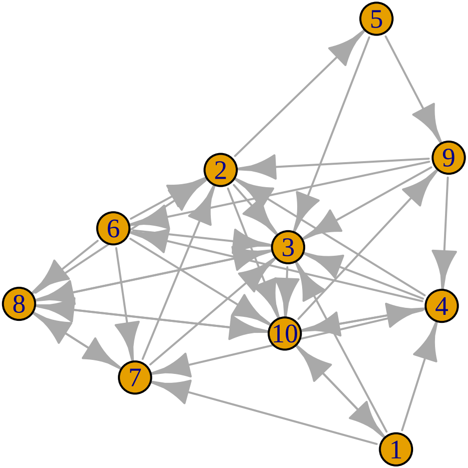
Undirected Graph

An undirected graph has no direction on the edges.



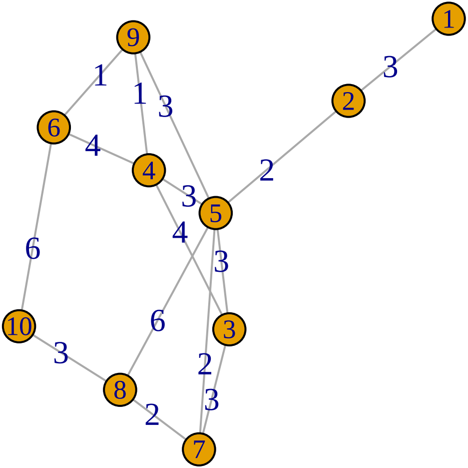
Directed Graph

An Directed graph has direction on the edges.



Weighted Graph

A Weighted Graph has weighted edges.



Formal Definition

A graph is a set of vertices, connected by edges. A graph is given by $G = (V, E)$, where V is the set of vertices:

$$V = \{v_1, v_2, \ldots, v_n\}$$

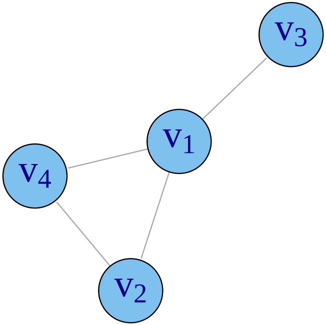
and E is the set of edges (pairs of vertices):

$$E = \{e_1, e_2, \ldots, e_n\}, \quad e_i = \{v_i, v_j\}, \quad v_i, v_j \in V$$

We define this graph as:

- $V = \{v_1, v_2, v_3, v_4\}$
- $E = \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_2, v_4\}\}$

The number of vertices is $|V|$ (the size of set V). The number of edges is $|E|$ (the size of set E).



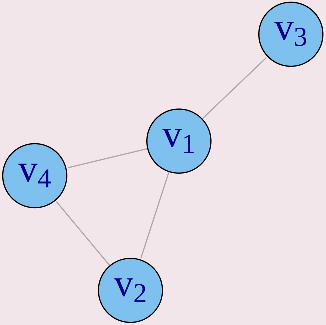
Representing Graphs using an Adjacency List

A graph is a set of vertices and edges. To store a graph, we do not need to store the image, we only need to store the vertices and edges. A simple method of representing a graph is an Adjacency List, listing each pair of vertices connected with edges.

Example Adjacency List

We define this graph as:

- v_1, v_2
- v_1, v_3
- v_1, v_4
- v_2, v_4



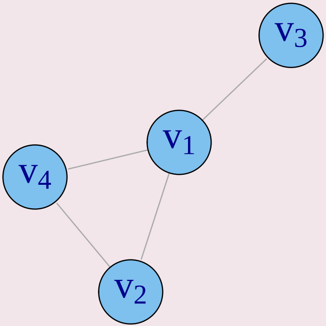
Representing Graphs using an Adjacency Matrix

An Adjacency Matrix A is a $|V| \times |V|$ matrix (where $|V|$ is the number of vertices). The element $a_{i,j}$ in the i row and jth column is 1 if there is an edge between the $i$the and jth vertice and 0 if there is no edge.

Example Adjacency Matrix

We define this graph as:

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$



Adjacency Matrix Problem

WESTERN SYDNEY UNIVERSITY

Problem

Write down the adjacency matrix for the following graph:

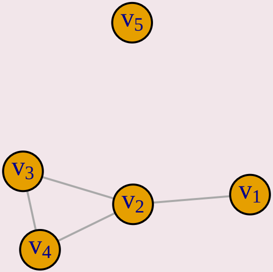


Figure: An undirected graph.

Outline

Subgraphs

A subgraph $H = (V_H, E_H)$ of the graph $G = (V_G, E_G)$ contains a subset of the vertices of G ($V_H \subset V_G$) and a subset of the interconnected edges ($E_H \subset E_G$).

All Subgraphs

There are 49 different subgraphs of the following graph. List all 49.

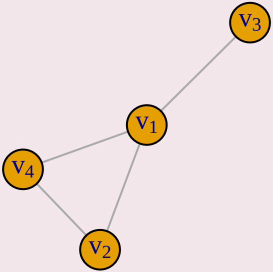


Figure: Another undirected graph.

Complete Graph (Clique)

A set of vertices where all vertices have edges between them, is called a complete graph, or a clique.

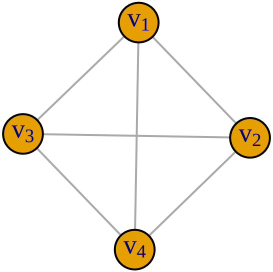


Figure: A complete graph (clique).

Egocentric Graphs

WESTERN SYDNEY UNIVERSITY

An egocentric graph is a subgraph that is centred around a chosen vertex.

A degree 1 egocentric graph is a subgraph that contains the chosen vertex, all connected vertices and the edges between them.

Degree 1 egocentric graph of v_1

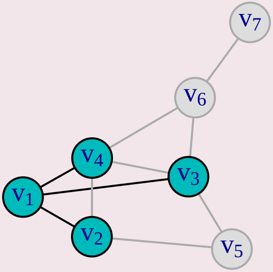


Figure: Degree 1 egocentric graph of v_1.

Egocentric Graphs

WESTERN SYDNEY UNIVERSITY

A degree 1.5 egocentric graph is a subgraph that contains the chosen vertex, all connected vertices and the edges between all vertices.

Degree 1.5 egocentric graph of v_1

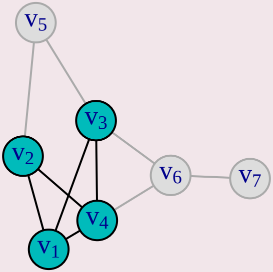


Figure: Degree 1.5 egocentric graph of v_1.

Egocentric Graphs

A degree 2 egocentric graph is a subgraph that contains the chosen vertex, all connected vertices and edges that are at most two hops from the chosen vertex.

Degree 2 egocentric graph of v_1

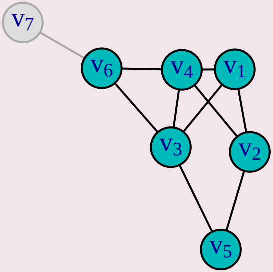
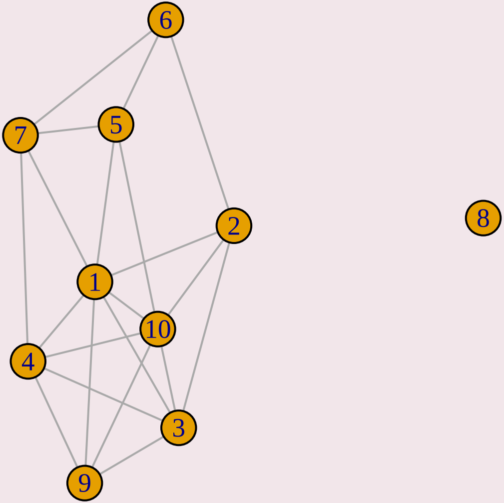


Figure: Degree 2 egocentric graph of v_1.

Extract the Graph

WESTERN SYDNEY UNIVERSITY

Find the degree 1 and 1.5 egocentric graph centred at vertex 1.



Outline

WESTERN SYDNEY UNIVERSITY

Vertex Degree

WESTERN SYDNEY UNIVERSITY

The degree of a vertex is measured as the number of edges that connect to the vertex.

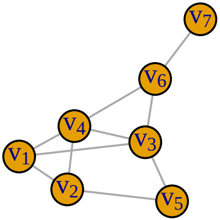


Figure: Degree of each vertex.

The degree of v_1 is 3. The degree of v_7 is 1.

Highest degree

Which vertex has the largest degree?

Degree Distribution

By examining the degree of all of the vertices, we can compute the degree distribution of the graph.

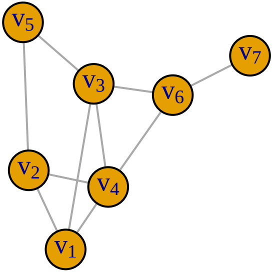


Figure: Compute the degree of each vertex to obtain the distribution.

Degree	0	1	2	3	4	5
Frequency						

Graph Density

WESTERN SYDNEY UNIVERSITY

A dense graph has many edges. A sparse graph has only a few edges. The density of a graph is the ratio of the edges that exist, compared to the possible edges.

Possible number of edges

Given $|V|$ vertices, the maximum number of edges is:

$$\sum_{i=1}^{|V|} i = |V|(|V| - 1)/2$$

Graph Density

The density D of graph $G = (V, E)$ is:

$$D = \frac{\text{Number of edges}}{\text{Possible number of edges}} = \frac{2|E|}{|V|(|V| - 1)}$$

Path

A path is a sequence of edges that are connected with vertices. Paths can contain the same edge multiple times. Therefore, we can make a path from a given start vertex and end vertex as long as we like.

Two paths from v_1 to v_6

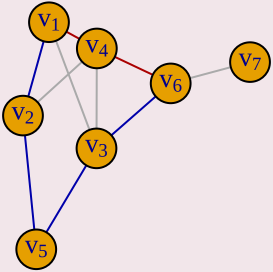


Figure: Two paths (one in red and one in blue).

Shortest Path

The shortest path between two vertices is the one that contains the least number of edges. Note the the shortest path may not be unique (there may be multiple paths that are just as short).

Two shortest paths from v_1 to v_6

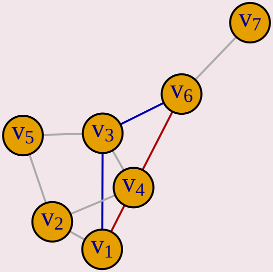


Figure: Two shortest paths from v_1 to v_6.

Graph Diameter

WESTERN SYDNEY UNIVERSITY

The diameter of a circle is the longest straight line that can fit in the circle.

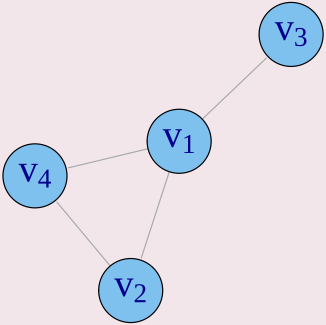
The diameter of a graph is the longest shortest path in the graph.

To find the diameter, we must find the shortest paths between all pairs of vertices. The diameter is the length of the longest of these paths.

All shortest paths

- v_1 to v_2: 1
- v_1 to v_3: 1
- v_1 to v_4: 1
- v_2 to v_3: 2
- v_2 to v_4: 1
- v_3 to v_4: 2

The diameter of the graph is 2.



Graph Parameter Problem

WESTERN SYDNEY UNIVERSITY

Problem

Compute the density and diameter of the following graph.

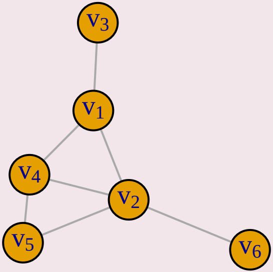


Figure: Density and diameter.

Outline

Requirements

Requirements for a centrality score:

1. the centrality score of a vertex correlates with the importance or popularity of the vertex,
2. it is computable for large graphs in reasonable time, and
3. it is difficult for individuals to manipulate for personal gain.

Measures of Centrality

A measure of centrality directs us to the centre of an object, therefore the location of the centre depends on how we define the centre.

For continuous numerical data, we can compute the centre as the mean, median or mode, depending on how we define the centre. For graphs there are also many measures of centrality.

There is usually no clear centre of a graph, therefore centrality methods provide a centrality score to all vertices, where the greater the score, the more likely the centre exists at that associated vertex.

Degree Centrality

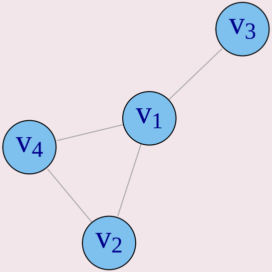
The simplest measure of centrality is *Degree Centrality*. The centrality score for each vertex is equal to the number of edges connected to it. The reasoning for this measure is, the greater the degree of a vertex, the more likely that information will flow through it.

$$C_{\text{degree}}(v) = \deg(v)$$

Example

- $\deg(v_1) = 3$
- $\deg(v_2) = 2$
- $\deg(v_3) = 1$
- $\deg(v_4) = 2$

The centre of the graph is v_1.



Degree Centrality Problem

Problem

Find the most central vertex, using degree centrality.

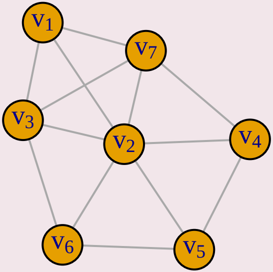


Figure: Compute degree centrality.

Closeness Centrality

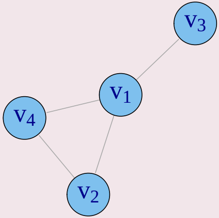
A vertex should be considered the centre if we can reach all other vertices from it with the shortest number of hops.

If we define $d(v_i, v_j)$ as the number of hops along the shortest path from v_i to v_j, then we define closeness centrality as:

$$C_{\text{closeness}}(v) = \sum_{u \in V} d(v, u)$$

Example

- $C_{\text{closeness}}(v_1) = 1 + 1 + 1 = 3$
- $C_{\text{closeness}}(v_2) = 1 + 2 + 1 = 4$
- $C_{\text{closeness}}(v_3) = 1 + 2 + 2 = 5$
- $C_{\text{closeness}}(v_4) = 1 + 1 + 2 = 4$



The centre of the graph is v_1.

Closeness Centrality Problem

Problem

Find the most central vertex, using closeness centrality.

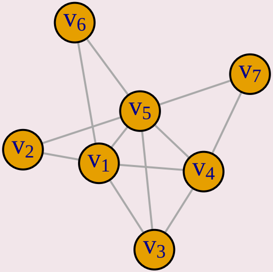


Figure: Compute closeness centrality.

Between Centrality

Centrality can be measured by the usefulness of the vertex. If the shortest path between vertices u_i and u_j contain vertex v, then the betweenness centrality of v is considered increase.

If we define sp_{u_i,u_j} as the number of shortest paths from vertices u_i to u_j and $\mathrm{sp}_{u_i,u_j}(v)$ as the number of shortest paths from vertices u_i to u_j that pass through v, then the betweenness centrality of v is:

$$C_{\text{betweenness}}(v) = \sum_{u_i,u_j \in V, u_i \neq u_j} \frac{\mathrm{sp}_{u_i,u_j}(v)}{\mathrm{sp}_{u_i,u_j}}$$

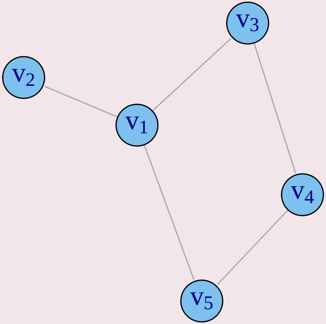
Note that a shortest path containing v at its start or end cannot pass though v. Therefore, we can ignore all paths that begin with v and end with v.

Between Centrality Example

Example

Shortest paths:

- $v_1 : v_2$
- $v_1 : v_3$
- $v_1 : v_3 : v_4, v_1 : v_5 : v_4$
- $v_1 : v_5$
- $v_2 : v_1 : v_3$
- $v_2 : v_1 : v_3 : v_4, v_2 : v_1 : v_5 : v_4$
- $v_2 : v_1 : v_5$
- $v_3 : v_4$
- $v_3 : v_1 : v_5, v_3 : v_4 : v_5$
- $v_4 : v_5$



$$C_{\text{betweenness}}(v_1) = 1/1 + 2/2 + 1/1 + 0/1 + 1/2 + 0/1 = 3.5$$

Between Centrality problem

WESTERN SYDNEY UNIVERSITY

Problem

Find the most central vertex, using between centrality.

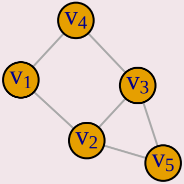


Figure: Compute between centrality.

Eigenvalue Centrality

Eigenvalue centrality computes the centre by stating that the centrality of a vertex is proportional to the sum of the centrality of its connected vertices.

$$C_{\text{eigen}}(v_i) \propto \sum_{v_j \in L(v_i)} C_{\text{eigen}}(v_j)$$

If we have the adjacency matrix A with elements $a_{i,j}$, the above becomes:

$$\lambda C_{\text{eigen}}(v_i) = \sum_{v_j \in G} a_{i,j} C_{\text{eigen}}(v_j)$$

or in vector notation:

$$\lambda \vec{c} = A \vec{c}$$

which is an eigenvalue problem. Note that λ is a constant that allows us to change the $\propto$ to $=$. The eigenvector $\vec{c}$ of the matrix A contains the centrality scores. Note that there may be many eigenvectors of A, but only one will contain all positive elements. We choose this as the solution.

Use of Centrality

Kevin Bacon is not the centre of the movie universe:

http://www.buzzfeed.com/tessastuart/
40-people-with-more-hollywood-connections-than-kevin-bacon

Outline

WESTERN SYDNEY UNIVERSITY

Using random graphs

Randomly generated models allow us to examine the behaviour of random processes. If we obtain a graph from a social network on one day, the graph might be different the next day. If we can find a random model that closely follows a random process, we gain insight into the random process.

Erdös-Renyi Graph

The simplest random graph is the Erdös-Renyi graph. The parameters of the random graph are the number of vertices n, and the probability of each edge existing p.

- Given n vertices, there are a possible $n(n-1)/2$ edges.
- For each possible edge, we sample a number from the Uniform distribution (between 0 and 1).
- If the random sample is less than p, the edge is added.

Example Erdös-Renyi Graphs

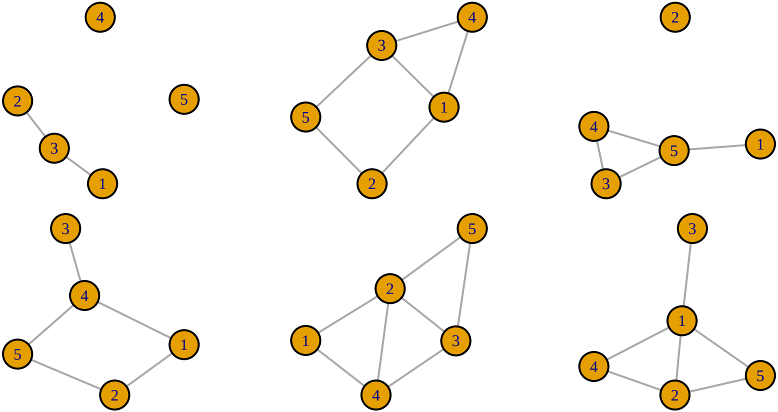


Figure: Erdös-Renyi Graphs with $n = 5, p = 0.5$.

Barabasi-Albert Graph

Barabasi-Albert random graph is generated in a way that reflects the growth of many evolving graphs such as the Web and social networks. The parameters are the number of vertices n, the number of edges to be added per step e, and the power k.

- Start with one vertex
- Place each vertex one at a time.
- When a vertex is placed it is randomly connected to e existing vertices.
- The probability of connecting to vertex v_i is:

$$P(v_i) = \frac{\deg(v_i)^k}{\sum_{v_j \in V} \deg(v_j)^k}$$

New vertices are more likely to added to vertices with a high degree.

Example Barabasi-Albert Graphs

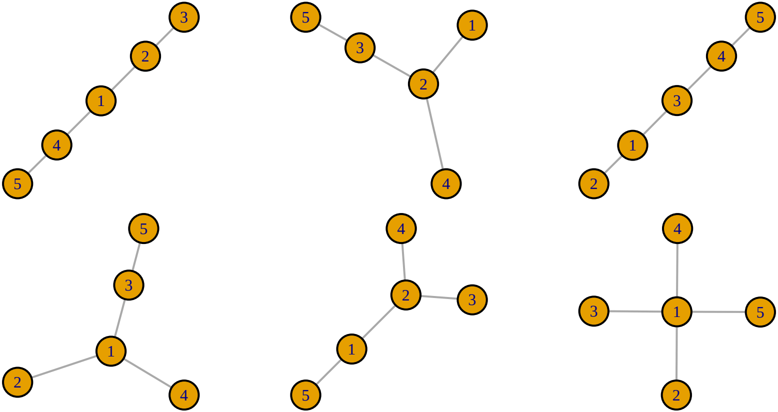


Figure: Barabasi-Albert Graphs with $n = 5, k = 1$.

Example Barabasi-Albert Graph

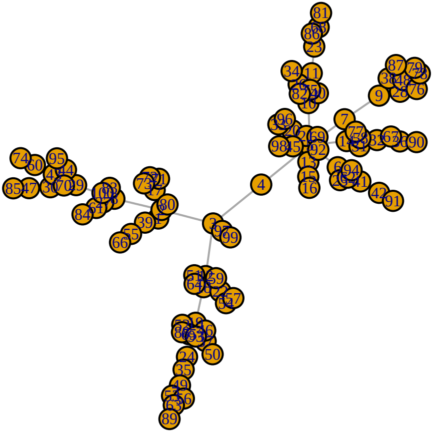


Figure: Barabasi-Albert Graphs with $n = 100, k = 1$.

Degree Distributions

BA model: Many vertices with low degree, few with high degree.

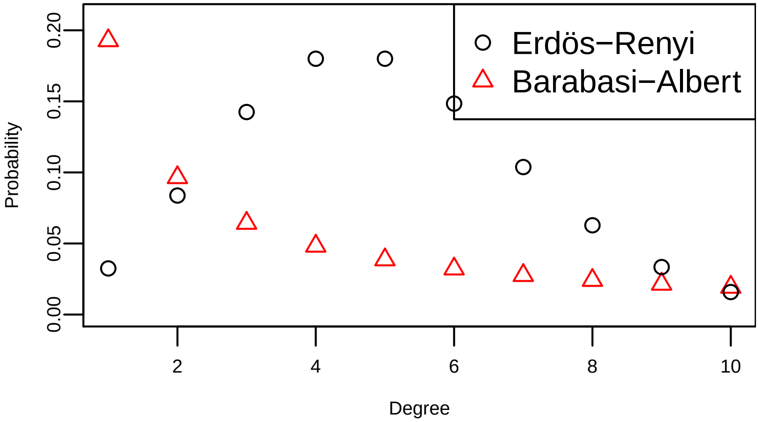


Figure: Distribution of Barabasi-Albert and Erdos-Renyi random graphs.

Properties of Social Networks

Social Networks have shown to follow the Barabasi-Albert model, having a power law distribution (exponentially decreasing) of vertex degree.

Social networks show a few vertices are hubs (have a high degree).

Social networks a robust: if vertices are removed at random, there are still paths between all vertices. But if carefully selected vertices (the hubs) are removed, the network will become disjoint.

Social networks form clusters and sub-clusters.

The Barabasi-Albert model is a good model for simulating social networks.

Summary

WESTERN SYDNEY UNIVERSITY

- Graph theory is essential for analysis of social networks.
- Graphs consist of vertices and edges. They can also be directed and weighted.
- There are various subgraphs that we can examine.
- There are various ways to measure centrality.
- We are able to generate random graphs using statistics from real graphs.
- Social networks resemble a Barabasi-Albert random graph.

Next Week

Graphs 2 - Link Analysis.