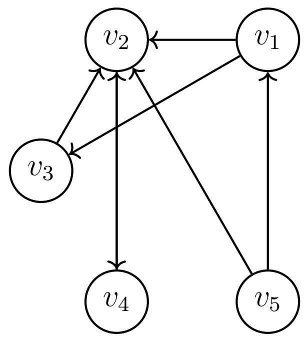
Graph theory

Author: Daniel Gigliotti

1 Graphs

You know what a graph is



We can describe this graph with an *adjacency matrix*:

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

If the graph is undirected, the adjacency matrix is symmetric (duh) and we can consider only the upper (or lower) triangle.
If the graph is directed, the adjacency matrix is not symmetric.

2 Density

The density L of a graph is the ratio between the number of edges L in the graph and the maximum possible number of edges L_{tot}

$$k = L/L_{tot}$$

- $L_{tot} = \frac{N(N-1)}{2}$ for *undirected* graphs (half adjacency matrix)
- $L_{tot} = N(N-1)$ for *directed* graphs (full adjacency matrix)

The density for the above graph is $L_{tot} = N(N-1) = 5(4) = 20$

3 Degree

The degree g of a node i ($i \in [1, N]$, N = number of nodes in the graph) is the number of edges connected to that node:

$$g_i = \sum_{j=0, j\neq i}^{N-1} a_{ij}$$

$$\text{with } g_i \in [0, N-1]$$

If the graph is directed:

- in-degree of a node i ($i \in [0, N-1]$) is the number of edges directed to node i
- out-degree of a node i ($i \in [0, N-1]$) is the number of edges originated from node i
- degree of a node i ($i \in [0, N-1]$) is the number of edges originated from and directed to node i

$$g_i = g_i^{in} + g_i^{out}$$

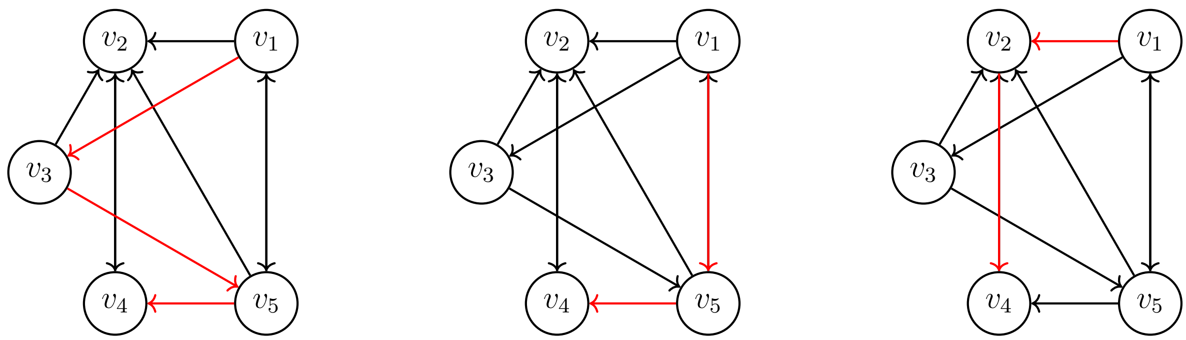
$$\text{with } g_i \in [0, 2(N-1)]$$

If we want to compute the degrees for each node:

out-degree	in-degree	degree
$g_1^{out} = 2$	$g_1^{in} = 1$	$g_1 = 3$
$g_2^{out} = 1$	$g_2^{in} = 4$	$g_2 = 5$
$g_3^{out} = 1$	$g_3^{in} = 1$	$g_3 = 2$
$g_4^{out} = 1$	$g_4^{in} = 1$	$g_4 = 2$
$g_5^{out} = 2$	$g_5^{in} = 0$	$g_5 = 2$

4 Path

A *path* is any possible sequence of edges linking two nodes. The *path length* is given by the number of edges from which the path is composed.



3 paths linking v_1 to v_4

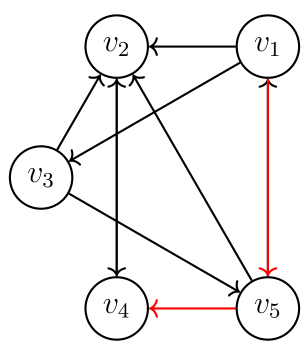
$$v_1 \to v_3 \to v_5 \to v_4 \ (3steps)$$

$$v_1 \to v_5 \to v_4 \ (2steps)$$

$$v_1 \to v_2 \to v_4 \ (2steps)$$

5 Distance

Distance $d(i, j)$ between nodes i and j is the *shortest path* between them



$$v_1 \to v_5 \to v_4 \ (2steps)$$
$$d(v_1, v_5) = 2 \to \ \textit{Shortest path between } v_1 \textit{ and } v_4$$

We can even build a *distance matrix*

$$D = \begin{bmatrix} 0 & 1 & 1 & 2 & 1 \\ \infty & 0 & \infty & 1 & \infty \\ 2 & 1 & 0 & 2 & 1 \\ \infty & 1 & \infty & 0 & \infty \\ 1 & 1 & 2 & 1 & 0 \end{bmatrix}$$

6 Global efficiency

The *global efficiency* E_g of a graph is the average of the reciprocal of the distances between any pair of nodes. It measures how efficiently the information is exchanged in the network.

$$E_g = \frac{1}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} \frac{1}{d_{ij}} \quad \text{for directed graphs}$$

$$E_g = \frac{2}{N(N-1)} \sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} \frac{1}{d_{ij}} \quad \text{for undirected graphs}$$

$$\text{with } E_g \in [0, 1]$$

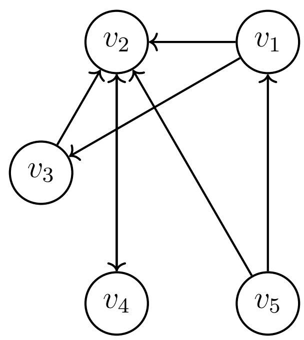
In general:

$$E_g = \frac{1}{L_{tot}} \sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} \frac{1}{d_{ij}}$$

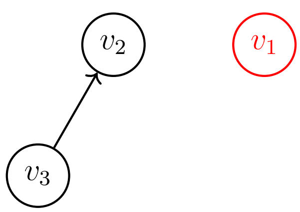
7 Local efficiency

Local efficiency measures the tendency of the network to form strongly connected subgroups. It is a measure of the robustness of the network to the removal of a node. For each node i we can extract a subnetwork S_i made of all the nodes directly connected to i excluding i itself and compute its global efficiency $E_g(S_i)$

Starting from this graph

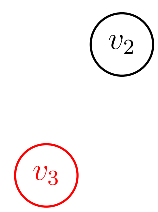


We will have:



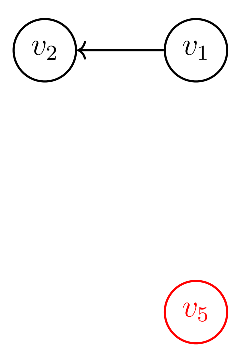
For v_1

For v_2



For v_3

For v_4



For v_5

$$D = \begin{bmatrix} 0 & \infty \\ 1 & 0 \end{bmatrix}$$

For v_1

$$D = \begin{bmatrix} 0 \end{bmatrix}$$

For v_2

$$D = \begin{bmatrix} 0 \end{bmatrix}$$

For v_3

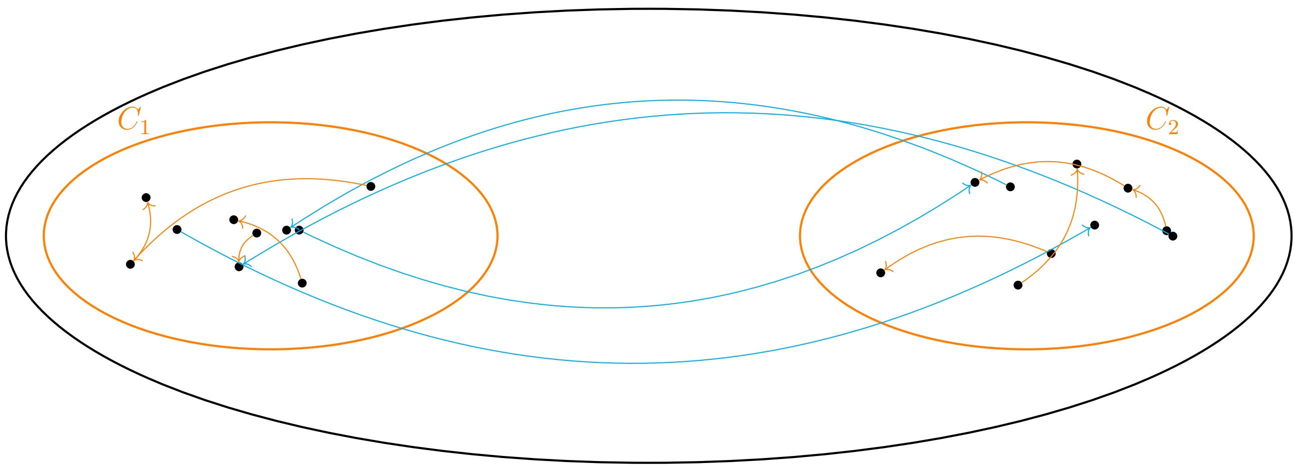
$$D = \begin{bmatrix} 0 \end{bmatrix}$$

For v_4

$$D = \begin{bmatrix} 0 & 1 \\ \infty & 0 \end{bmatrix}$$

For v_5

For some applications it may be interesting to measure how well a network can be divided into *communities*.



A subnetwork is a community if it shows an organized structure with respect to the entire network. Measures useful to determine if two or more subnetworks act as communities are the *divisibility* and *modularity*.

8 Divisibility

Divisibility D is a measure of the segregation between two communities. It focuses on inter-community links.

$$D = \frac{2L}{2L+\sum_{i,j=0}^{N-1} a_{ij}[1-\delta(C_i,C_j)]} \text{ for } \textit{undirected} \text{ graphs}$$

$$D = \frac{L}{L+\sum_{i,j=0}^{N-1} a_{ij}[1-\delta(C_i,C_j)]} \text{ for } \textit{directed} \text{ graphs}$$

$$D \in [\tfrac{1}{2}, 1]$$

$$\begin{cases} \delta(C_i, C_j) = 1 if C_i == C_j \\ \delta(C_i, C_j) = 0 otherwise \end{cases}$$

δ tells us if the two nodes belongs to the same community
L is the number of edges in the network

9 Modularity

Modularity Q measures the tendency of the subnetworks to form communities. It focuses on intra-community links.

$$Q = \frac{1}{2L} \sum_{i,j=0}^{N-1} (a_{ij} - \frac{g_i g_j}{2L}) \delta(C_i, C_j) \text{ for } \textit{undirected} \text{ graphs}$$

$$Q = \frac{1}{L} \sum_{i,j=0}^{N-1} (a_{ij} - \frac{g_i^{in} g_j^{out}}{L}) \delta(C_i, C_j) \text{ for } \textit{directed} \text{ graphs}$$

$$\begin{cases} \delta(C_i, C_j) = 1 if C_i == C_j \\ \delta(C_i, C_j) = 0 otherwise \end{cases}$$

δ tells us if the two nodes belongs to the same community
L is the number of edges in the network

For each pair of nodes belonging to the same community, modularity Q compares the existance (or not) of a link between them with the probability of having a connection between them in a random distribution of the links.

existance of the intra-community link

$$a_{ij} - \frac{g_i g_j}{2L}$$

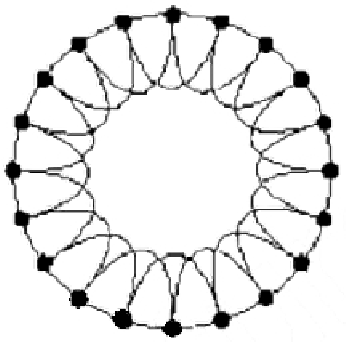
$$a_{ij} - \frac{g_i^{in} g_j^{out}}{L}$$

probability of a random link between the nodes

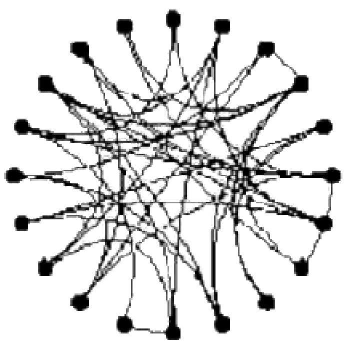
If there are more intra-community links than in a random division, Q > 0, otherwise Q $\leq$ 0

Integration	Segregation
high global efficiency	low global efficiency
low global efficiency	high global efficiency
low global efficiency	high global efficiency
low global efficiency	high global efficiency

10 Reference networks

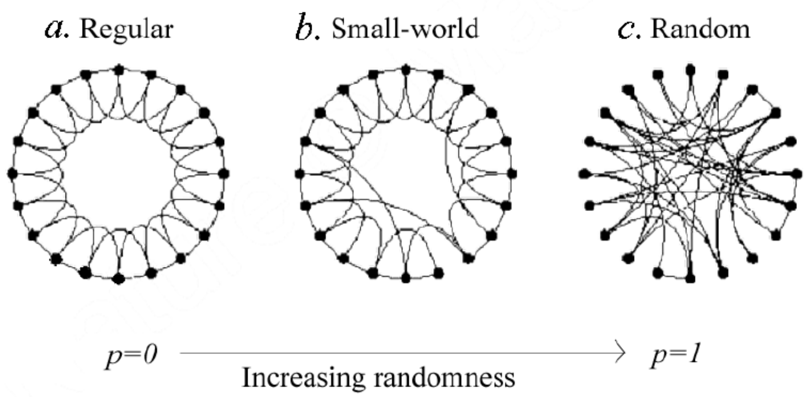


Regular networks: each node is linked to a small number of other nodes. The degree is the same for all nodes. Efficient communications between small groups (high local efficiency), unefficient communication at the entire network level (low global efficiency).



Random networks: each node is linked to the others randomly. There are no small groups with a strong internal communication (high global efficiency, low local efficiency).

Those two are ideal cases: real networks are neither like the random nor like the regular graphs.



$$E_g(Regular) < E_g(Small-world) < E_g(Random)$$
$$E_l(Random) < E_l(Small-world) < E_l(Regular)$$