Network 1

Giorgio Ricchiuti
www.grarchive.net
giorgio.ricchiuti@unifi.it

UNIVERSITÀ DEGLI STUDI FIRENZE

UNIVERSITÀ CATTOLICA del Sacro Cuore
CLE
Complexity Lab in Economics

CompEc

- Networks analysis enables the reconstruction of the linkages and the evolution of connections between different individuals/agents/firms.
- Specifically, the main effort has been to understand the basic mechanism of communication networks: internet, World Wide Web, e-mails network.
- The economic system is a natural net.

Sources

- Main textbook: M.E.J. Newman, *Networks: An Introduction*. Oxford University Press, 2010
- Other sources: Barabási, Vespignani, Dorogovtsev, Caldarelli...
- On Economic/Social networks: Jackson; Easley and Kleinberg; Vega-Redondo.
- Software packages: `igraph` (R), `networkx` (Python)
- Visualization software: Pajek, Gephi, Cytoscape...

Applications in Economics

In the empirical literature network analysis is mostly used:

- to study the structure of the interbank market (Iori et al., 2008);
- bank-firm credit market (De Masi et al., 2011; Battiston et al. 2007);
- financial market investments (Garlaschelli et al., 2005);
- the world trade network (De Benedictis and Tajoli, 2011);
- the world trade web (Fagiolo et al., 2011);
- on FDI at global level: Wall et al. (2011), Alfaro and Chen (2014);
- on Italian FDI (De Masi et al., 2013) and French data (Joyez, 2017).

Definition

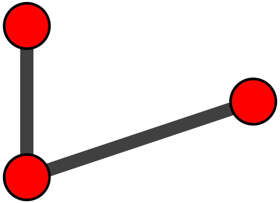
'A Network is, in its simplest form, a collection of points joined together in pairs by lines.'

- A *point* is a **node** or **vertex**;
- A *line* is a **link** or, better, an **edge**.

Formally, a **network** (or **graph**) G is a pair of sets $G = (V, E)$, where:

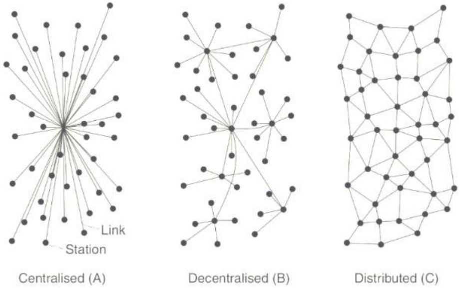
- V is the set of **vertices** (nodes); $|V| = n$ is the **order** of the graph.
- E is the set of **edges** (links); $|E| = m$ is the **size** of the graph.

A simple example



Centralized, Decentralized, Distributed

1. **Centralized**: built around a single master node (e.g. YouTube, bank accounts).
2. **Decentralized**: multiple master nodes, each interacting independently with others (e.g. Facebook).
3. **Distributed**: composed of equal, interconnected nodes (e.g. Internet).



Technological Networks

A good example is Internet:

- Nodes are computers
- Edges are data connections between computers (optical fiber cables or telephone lines)
- ...the function of Internet is to transport data between computers, which it does by dividing the data into separate packets and shipping them from *node to node* across the net.
- The network structure will affect how efficiently it performs this function... therefore it is worth studying.

Networks of Information

A more abstract class that represents the network structure of bodies of information. Classic example: the WWW.

- Nodes are web pages
- Hyperlinks between web pages, we click on to navigate from one page to another
- A hyperlink is a software construct; there is no physical structure. . . the link is just an address.
- The link structure of WWW reflects human knowledge.

Another information network: citation networks (citations between academic journal articles).

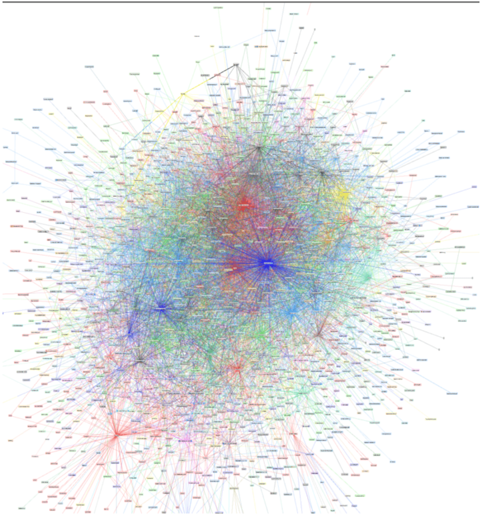


Figure: WWW network

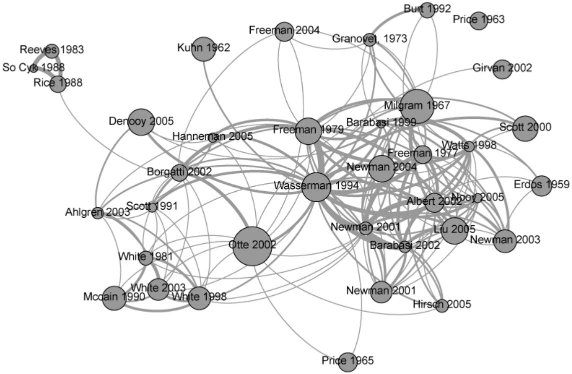
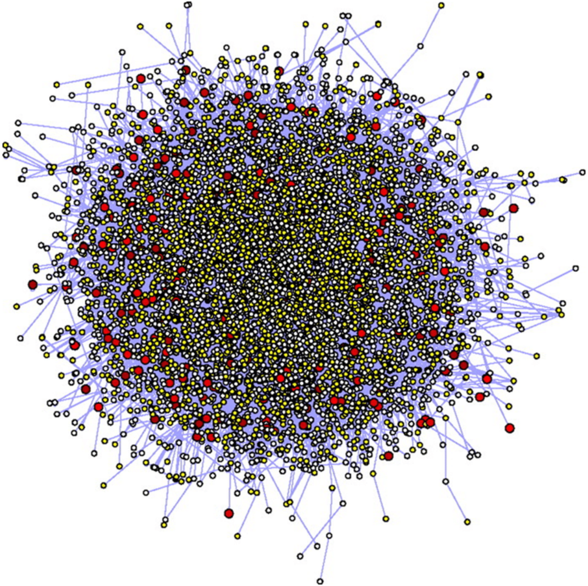


Figure: Citation network

Social Networks

Not just Facebook, Twitter or Instagram. Within the scientific literature the term encompasses any network in which:

- Nodes are people (or groups of people, firms, teams. . .)
- The edges refer to social connections of some kind: friendship, communication, or collaboration.
- Sociologists were the first to study these. . .



Biological Networks

Networks occur in a range of different settings in biology:

- **Neural**: connections between neurons in the brain
- **Food web**: ecological network in which nodes are species and edges represent predator-prey relationships
- **Biochemical**: protein-protein interaction network. . .

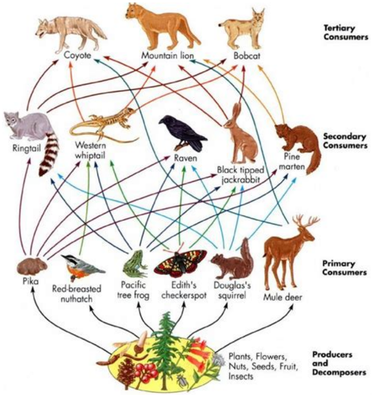


Figure: Food Web 1

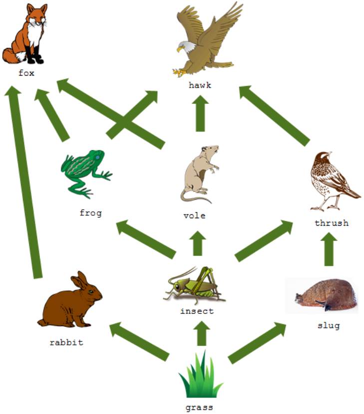


Figure: Food Web 2

Adjacency Matrix

A network is represented mathematically by an **adjacency matrix**.

The element a_{ij} indicates that a link exists between nodes i and j:

$$a_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The **degree** of node i is the number of its links:

$$k_i = \sum_j a_{ij} \tag{2}$$

The **density** of the network:

- Undirected: $c = \frac{2L}{n(n-1)}$ Directed: $c = \frac{L}{n(n-1)}$

where $L = \sum_{i<j} a_{ij}$ is the total number of edges.

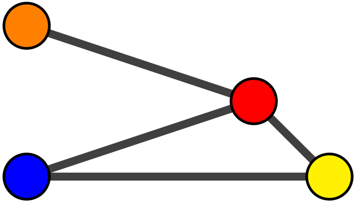
Self-edges and Multiedges

- There can be more than one edge between the same pair of vertices: **multiedges**.
- There can be edges connecting a vertex to itself: **self-edges**.

A network with neither self-edges nor multiedges is called a **simple network**.
A network with multiedges is called a **multigraph**.

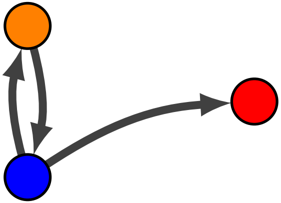
- A multiedge is represented by setting A_{ij} equal to the multiplicity of the edge.
- A self-edge from i to itself is represented by setting the diagonal element $A_{ii} = 2$.

Undirected Network



If E is a set of **non-ordered** pairs of distinct elements in V: **undirected network**.
The adjacency matrix is symmetric: $A = A^T$.

Directed Network



If E is a set of **ordered** pairs of distinct elements in V: **directed network**.
A link from i to j is different from a link from j to i.

Degree splits into: **in-degree** $k_i^{in} = \sum_j a_{ji}$ and **out-degree** $k_i^{out} = \sum_j a_{ij}$.

Weighted Networks

If a network is associated with a function $w : V \to \mathbb{R}^+$ we have a **weighted network**: w_{ij} is the weight of link i–j.

The weights can represent:

- quantities of flows (trade volumes, financial exposures...)
- strength of relationships (frequency of communication...)

For weighted networks the **strength** of a node is:

$$s_i = \sum_j w_{ij} \tag{3}$$

Similarly one defines in-strength and out-strength for directed weighted networks.

Paths and Distance

- A **path** is any sequence of vertices such that every consecutive pair is connected by an edge.
- The **length** of a path is the number of edges traversed.
- The number of paths of length r from i to j is $(A^r)_{ij}$.
- The total number of **cycles** of length r: $L_r = \mathrm{Tr}[A^r] = \sum_i \phi_i^r$, where ϕ_i are eigenvalues of A.

The **distance** d_{ij} between two vertices i, j is the *shortest* number of edges to go from i to j:

$$d_{ij} = \min \left\{ \sum_{k,l \in \mathcal{P}_{ij}} a_{kl} \right\} \qquad (4)$$

where $\mathcal{P}_{ij}$ is a path connecting i and j.

The **diameter** of a graph is the maximum of all pairwise distances.

Bipartite Networks

Let the two kinds of nodes be A and B. We can study:

- the network G_{A+B} with the full set of nodes;
- the **projected** networks G_A and G_B with only nodes of kind A or B.

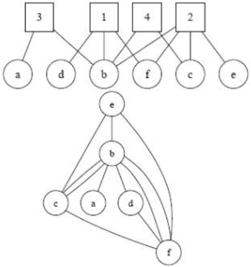


Figure: Top: bipartite graph (squares = countries; circles = parent firms). Bottom: projected graph on firm space.

What Can We Learn from Networks?

- Networks capture the pattern of interactions between the parts of a system.
- A net is a simplified representation that reduces a system to an abstract topology.
- It creates a bridge between empirical data and a large toolkit of powerful analysis techniques.

Given the topology we can calculate local, meso, and large-scale topological measurements:

A) Local:

- 1st order: **degree**
- 2nd order: avg. neighbor degree
- 3rd order: **clustering**
- 4th order: squared clustering

B) Higher order:

- betweenness
- closeness
- eigenvector centrality
- eccentricity

Centrality: the Big Picture

What can we do with networks?

The **centrality** of a node measures its **importance**:

- how influential is a person in a social network?
- how critical is an element in an infrastructure network?
- what is the disease-spreading capacity of an individual?
- what is the most systemically important financial institution?

- **Degree centrality** $\rightarrow$ **POPULARITY**
- **Betweenness centrality** $\rightarrow$ **BRIDGE**
- **Closeness centrality** $\rightarrow$ **CENTRALNESS**
- **Eigenvector centrality** $\rightarrow$ **INFLUENCE**

Degree and Degree Centrality

- The **degree** of node i is the number of its links:

$$k_i = \sum_j a_{ij}$$

- The **degree centrality** normalises by the maximum possible degree:

$$dc_i = \frac{k_i}{N-1} \quad (5)$$

where N is the total number of nodes.

Average Neighbor Degree and Assortativity

Starting from degree, we can calculate the **average neighbor degree**, which generally depends on the degree of the node considered (Caldarelli, 2007).

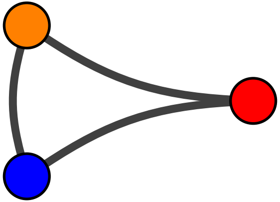
A high average neighbor degree means that the node is linked to highly-connected nodes.

This indicator also reveals whether a network is **assortative**:

- if the average neighbor degree *increases* with node degree: **assortative**
- if it *decreases*: **disassortative**

Clustering Coefficient

The **clustering coefficient** measures the density of connections around a vertex — the proportion of a node's neighbors that are connected to one another.



Betweenness Centrality

The **betweenness centrality** is given by the number of times that node i lies on a shortest path between two other nodes j and l:

$$b_i = \sum_{\substack{j,l=1 \\ i \neq j \neq l}}^{N} \frac{d_{jl}(i)}{d_{jl}} \tag{6}$$

where d_{jl} is the total number of shortest paths from j to l and $d_{jl}(i)$ is the subset passing through i.

Interpretation: a node with high betweenness acts as a **bridge** — removing it disconnects large parts of the network.

Closeness Centrality

The **closeness centrality** is the reciprocal of the average distance from node i to all other nodes:

$$cl_i = \frac{N-1}{\sum_j d_{ij}} = \frac{1}{\bar{d}_i} \tag{7}$$

In order to be a hub, a node should not be very distant from all the others.

Interpretation: how quickly can information starting from i reach the rest of the network?

Eigenvector Centrality

The **eigenvector centrality** measures the importance of a node based on the importance of its neighbors. In vector notation, it solves $W \cdot c = \lambda c$, i.e.:

$$c_i = \frac{1}{\lambda} \sum_j W_{ij} c_j \tag{8}$$

where λ is the largest eigenvalue of W.

It is intrinsically based on the **spectral properties** of the adjacency matrix.

Interpretation: being connected to well-connected nodes matters — this is the logic behind Google's PageRank.

Eccentricity

The **eccentricity** of a node is the inverse of its maximum distance from any other node:

$$e_i = \frac{1}{\max_{\forall j \in N} d_{ij}} \tag{9}$$

Related concepts:

- **Center**: set of nodes with minimum eccentricity.
- **Diameter**: maximum distance between any pair of vertices.
- **Radius**: minimum among all maximum distances, $r(G)$.