Introduction to complex networks

Gianluca Campanella

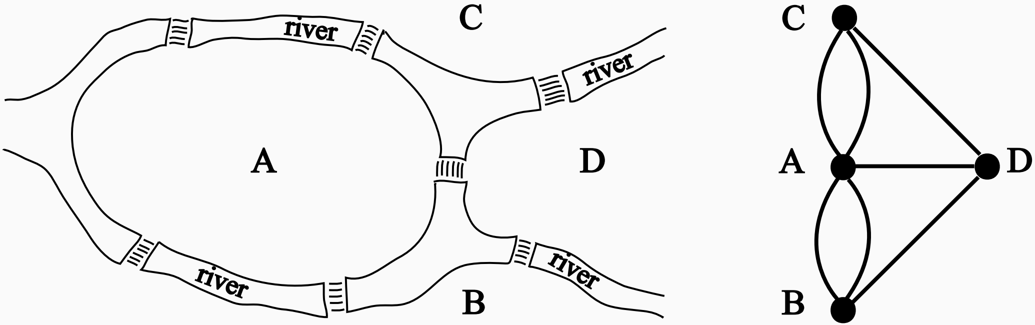
Contents

Definitions

Graph theory

1736 Euler writes the first paper

1959 Erdős and Rényi introduce probabilistic methods

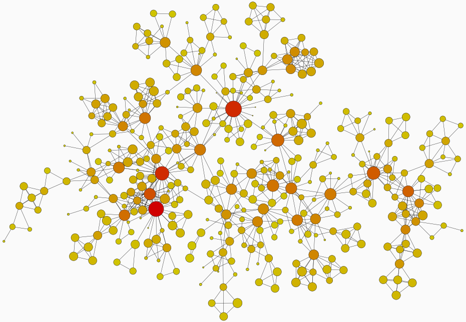


From Junker and Schreiber (2008)

Network theory

1970s Sociologists ('social network analysis')

2000s Physicists and Computer Scientists



Graphs and their representation

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of:

1. a set of **vertices** $\mathcal{V} = \{v_1, \ldots, v_n\}$
2. a binary relation $\mathcal{E} \subseteq \mathcal{V}^2$ that represents **edges**

Graphs and their representation

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of:

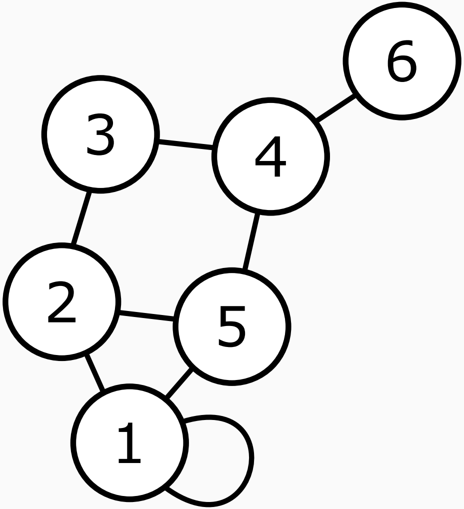
1. a set of **vertices** $\mathcal{V} = \{v_1, \dots, v_n\}$
2. a binary relation $\mathcal{E} \subseteq \mathcal{V}^2$ that represents **edges**

Alternative representation of $\mathcal{E}$

In terms of the **adjacency matrix** $\mathbf{A} \in \{0, 1\}^{n \times n}$, with elements

$$a_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

Example



$$\mathcal{V} = \{v_1, \ldots, v_6\}$$

$$\mathcal{E} = \{(v_1, v_1), (v_1, v_2), \ldots\}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Weights

- Assigned to elements of $\mathcal{E}$
- Represent the **strength** of the relationship

Examples

- Signed edges: $\mathbf{A} \in \{-1, 0, 1\}^{n \times n}$
- Weights on the unit interval: $\mathbf{A} \in [0, 1]^{n \times n}$
- Weights on the positive real line: $\mathbf{A} \in [0, \infty)^{n \times n}$

Properties of $\mathcal{E}$

Irreflexivity

$$\forall\, v_i \in \mathcal{V}.\, (v_i, v_i) \notin \mathcal{E}$$

$\rightarrow$ No ‘self-loops’ allowed

Symmetry

$$\forall\, v_i, v_j \in \mathcal{V}.\, (v_i, v_j) \in \mathcal{E} \Rightarrow (v_j, v_i) \in \mathcal{E}$$

$\rightarrow$ Direction of the relationship does not matter ($\mathcal{G}$ is **undirected**)

$\rightarrow$ $\mathbf{A} = \mathbf{A}^\top$ is also symmetric

Properties of vertices

Neighbourhood

$$\mathcal{N}(v_i) = \{v_j : (v_i, v_j) \in \mathcal{E}\} \subseteq \mathcal{V}$$

$\rightarrow$ **Set** of vertices to which v_i is connected by edges

Degree

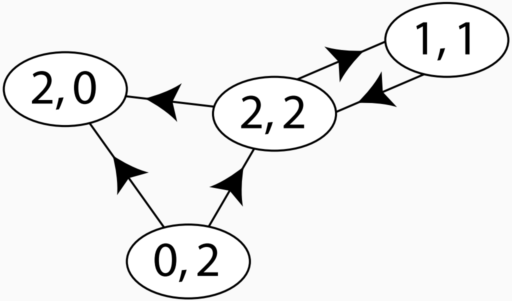
$$k_i = k(v_i) = |\mathcal{N}(v_i)|$$

$\rightarrow$ **Number** of vertices to which v_i is connected by edges

In- and out-degree

$$k_i^{\text{in}} = |\{(v_j, v_i) \mid (v_j, v_i) \in \mathcal{E}\}|$$

$$k_i^{\text{out}} = |\{(v_i, v_j) \mid (v_i, v_j) \in \mathcal{E}\}|$$



Paths

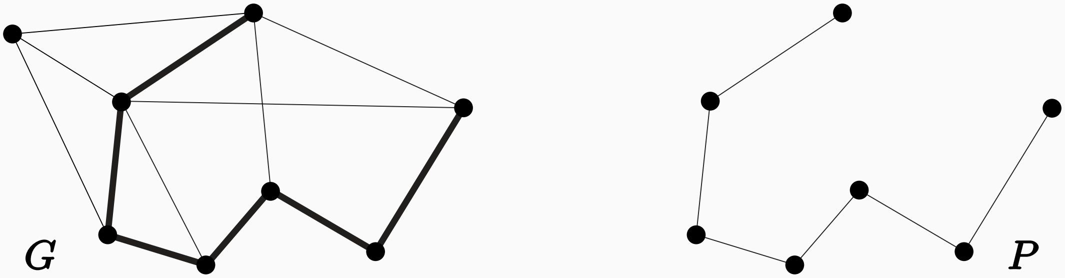
A **path** on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph $\mathcal{P} = (\mathcal{V}_\mathcal{P}, \mathcal{E}_\mathcal{P})$ with:

1. $\mathcal{V}_\mathcal{P} = \{v_{(0)}, \ldots, v_{(l)}\} \subseteq \mathcal{V}$
2. $\mathcal{E}_\mathcal{P} = \{(v_{(0)}, v_{(1)}), \ldots, (v_{(l-1)}, v_{(l)})\} \subseteq \mathcal{E}$

Ends Vertices $v_{(0)}$ and $v_{(l)}$

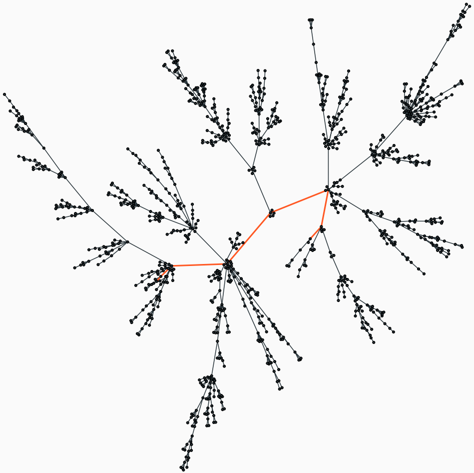
Length Number of edges $|\mathcal{E}_\mathcal{P}| = l$

Paths



From Diestel (2010)

Shortest paths



Local properties

Which are the most important
(or ‘central’) vertices?

Degree centrality
How many connections does a vertex have to other vertices?

Which are the most important
(or ‘central’) vertices?

Closeness centrality
How ‘far’ (in terms of shortest paths) is a vertex with respect to all other vertices?

Which are the most important (or ‘central’) vertices?

Betweenness centrality

How many times does a vertex act as a ‘bridge’ along the shortest paths between all pairs of vertices?

Centrality

Which are the most important
(or ‘central’) vertices?

Eigenvector centrality (PageRank)
How many times do we stumble upon a vertex while randomly walking on the graph?

Local clustering coefficient

$$C_i = C(v_i) = \frac{|\{(v_j, v_k) : \{(v_i, v_j), (v_i, v_k), (v_j, v_k)\} \subseteq \mathcal{E}\}|}{k_i(k_i - 1)/2} \in [0, 1]$$

- Probability that a pair of neighbours of v_i are connected
- Empirically: $k_i \nearrow$, $C_i \searrow$

$\rightarrow$ Can be used to identify 'structural holes'

Large-scale structure

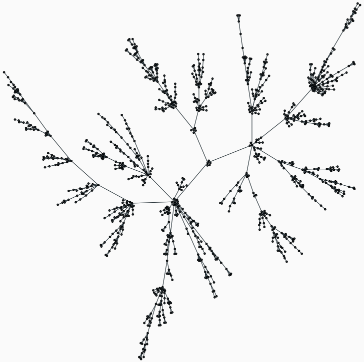
‘Small-world’ effect

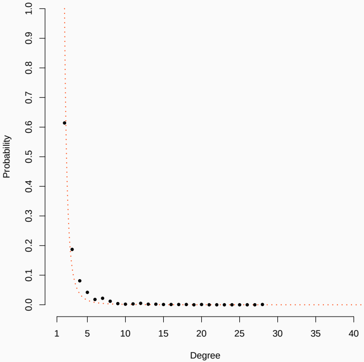
‘...Milgram’s letter-passing experiment in the 1960s, in which people were asked to get a letter from an initial holder to a distant target person by passing it from acquaintance to acquaintance through the social network. The letters that made it to the target did so in a remarkably small number of steps, around six on average.’

From Newman (2010)

Degree distribution

$$\Pr[K = k] = \Pr[\text{'a randomly selected vertex } v_i \in \mathcal{V} \text{ has degree } k \geq 0\text{'}]$$





Power laws

$$\ln y = -\alpha \ln x + c$$

$$y = Cx^{-\alpha}, \quad C = e^c$$

- Sizes of city populations and wars
- Frequency of use of words in human languages
- Occurrence of personal names in most cultures
- Number of papers scientists write
- Number of hits on Web pages
- Sales of almost every branded commodity
- Number of species in biological taxa

Scale-free networks

- $\Pr[K = k] \propto k^{-\alpha}$ (at least asymptotically), usually $2 \leq \alpha \leq 3$
- **Self-similar** (scale invariance)
- Presence of **hubs** (Milgram's 'sociometric superstars')
  - $\rightarrow$ Fault tolerance
  - $\rightarrow$ 'Small-world' effect

Scale-free networks

Examples

- Social and collaboration networks
- Many kinds of communication networks (e.g. the Internet)
- **Many biological networks**

Models of network formation

What is complexity?

Complex

Many interacting parts, not necessarily ‘difficult’

Complicated

Difficult, not necessarily ‘complex’

What is complexity?

Via *Wikimedia Commons*

Self-organisation

Spontaneous emergence of global structure out of local interactions

- Robust to damage and perturbations
  - Not controlled by (internal or external) agents
  - Collective, distributed process
- Outcome is not arbitrary, but ‘prefers’ certain situations
  $\rightarrow$ Natural selection

Self-organisation

Examples

- Protein folding
- Pattern formation and morphogenesis
- Social structures and herd behaviour

Topology is crucial to understanding stochastic processes on networks

Models of network formation

Random

Prefential attachment

Erdős–Rényi model

Algorithm

Construct a graph with n vertices and include each edge with probability p (independently from every other edge)

Erdős–Rényi model

Algorithm

Construct a graph with n vertices and include each edge with probability p (independently from every other edge)

Degree distribution

$$\Pr[K = k] = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Barabási–Albert model

Algorithm

- Start with an initial connected graph with n vertices
- Add a new vertex and connect it to $n^\star \leq n$ existing vertices with probability proportional to their degree
  $\rightarrow$ **Preferential attachment**

Barabási–Albert model

Algorithm

- Start with an initial connected graph with n vertices
- Add a new vertex and connect it to $n^\star \leq n$ existing vertices with probability proportional to their degree
  $\rightarrow$ **Preferential attachment**

Degree distribution

$$\Pr[K = k] \propto k^{-3}$$

What about ‘–omics’?

- Very few longitudinal datasets
  $\rightarrow$ Difficult to understand evolution
- Few cross-platform studies
  $\rightarrow$ Can understand interactions between complexity levels
- Many single-platform studies
  $\rightarrow$ Can characterise “co-” networks