# Network Science Cheatsheet

Made by
Remy Cazabet

# 1 Network basics

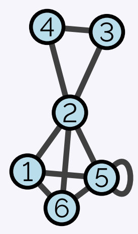
## Networks: Graph notation

Graph notation : $G = (V, E)$

| | |
|---|---|
| $V$ | set of vertices/nodes. |
| $E$ | set of edges/links. |
| $u \in V$ | a node. |
| $(u, v) \in E$ | an edge. |

## Network - Graph notation

**Graph**



**Graph notation**

$$G = (V, E)$$
$$V = \{1, 2, 3, 4, 5, 6\}$$
$$E = \{(1, 2), (1, 6), (1, 5), (2, 4), (2, 3), (2, 5), (2, 6), (6, 5), (5, 5), (4, 3)\}$$

## Types of networks

**Simple graph**: Edges can only exist or not exist between each pair of node.
**Directed graph**: Edges have a direction: $(u, v) \in V$ does not imply $(v, u) \in V$
**Weighted graph**: A weight is associated to every edge.

Other types of graphs (multigraphs, multipartite, hypergraphs, etc.) are introduced in sheet **??**

## Counting nodes and edges

| | |
|---|---|
| $N/n$ | **size**: number of nodes $\|V\|$. |
| $L/m$ | number of edges $\|E\|$ |
| $L_{max}$ | Maximum number of links |

Undirected network: $\binom{N}{2} = N(N-1)/2$

Directed network: $\binom{N}{2} = N(N-1)$

## Node-Edge description

| | |
|---|---|
| $N_u$ | **Neighbourhood** of $u$, nodes sharing a link with $u$. |
| $k_u$ | **Degree** of $u$, number of neighbors $\|N_u\|$. |
| $N_u^{out}$ | **Successors** of $u$, nodes such as $(u, v) \in E$ in a directed graph |
| $N_u^{in}$ | **Predecessors** of $u$, nodes such as $(v, u) \in E$ in a directed graph |
| $k_u^{out}$ | **Out-degree** of $u$, number of outgoing edges $\|N_u^{out}\|$. |
| $k_u^{in}$ | **In-degree** of $u$, number of incoming edges $\|N_u^{in}\|$ |
| $w_{u,v}$ | **Weight** of edge $(u, v)$. |
| $s_u$ | **Strength** of $u$, sum of weights of adjacent edges, $s_u = \sum_v w_{uv}$. |

## Network descriptors - Nodes/Edges

$\langle k \rangle$ **Average degree**: Real networks are sparse, i.e., typically $\langle k \rangle \ll n$. Increases slowly with network size, e.g., $d \sim \log(m)$

$$\langle k \rangle = \frac{2m}{n}$$

$d/d(G)$ **Density**: Fraction of pairs of nodes connected by an edge in $G$.

$$d = L/L_{\max}$$

## Paths - Walks - Distance

**Walk**: Sequences of adjacent edges or nodes (e.g., **1.2.1.6.5** is a valid walk)
**Path**: a walk in which each node is distinct.
**Path length**: number of edges encountered in a path
**Weighted Path length**: Sum of the weights of edges on a path
**Shortest path**: The shortest path between nodes $u, v$ is a path of minimal *path length*. Often it is not unique.
**Weighted Shortest path**: path of minimal *weighted path length*.
$\ell_{u,v}$**: Distance**: The distance between nodes $u, v$ is the length of the shortest path

## Network descriptors - Paths

| | |
|---|---|
| $\ell_{\max}$ | **Diameter**: maximum *distance* between any pair of nodes. |
| $\langle \ell \rangle$ | **Average distance**: |

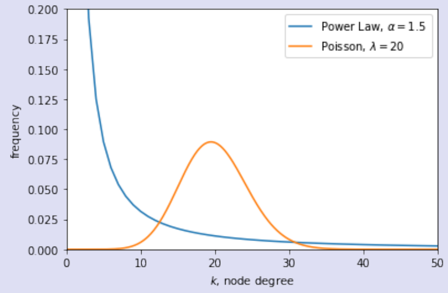
$$\langle \ell \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

## Degree distribution

The degree distribution is considered an important network property. They can follow two typical distributions:

- **Bell-curved** shaped (Normal/Poisson/Binomial)
- **Scale-free**, also called *long-tail* or *Power-law*

A Bell-curved distribution has a *typical scale*: as human height, it is centered on an average value. A Scale-free distribution has no typical scale: as human wealth, its average value is not representative, low values (degrees) are the most frequent, while a few very large values can be found (hubs, large degree nodes).




More details later.

## Subgraphs

**subgraph** $H(W)$: subset of nodes $W$ of a graph $G = (V, E)$ and edges connecting them in $G$, i.e., subgraph $H(W) = (W, E'), W \subset V, (u, v) \in E' \iff u, v \in W \wedge (u, v) \in E$
**Clique**: subgraph with $d = 1$
**Triangle**: clique of size 3
**Connected component**: a subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph
**Strongly Connected component**: In directed networks, a subgraph in which any two vertices are connected to each other by paths
**Weakly Connected component**: In directed networks, a subgraph in which any two vertices are connected to each other by paths if we disregard directions

## Triangles counting

$\delta_u$ - **Triads of** $u$**:** number of triangles containing node $u$
$\Delta$ - **Number of triangles in the graph** total number of triangles in the graph, $\Delta = \frac{1}{3}\sum_{u \in V} \delta_u$.

> Each **triangle** in the graph is counted as a **triad** once by each of its nodes.

$\delta_u^{\max}$ - **Triad potential of** $u$**:** maximum number of triangles that could exist around node $u$, given its degree: $\delta_u^{\max} = \tau(u) = \binom{k_i}{2}$
$\Delta^{\max}$ - **Triangle potential of G:** maximum number of triangles that could exist in the graph, given its degree distribution: $\Delta^{\max} = \frac{1}{3}\sum_{u \in V} \delta^{\max}(u)$

## Clustering Coefficents - Triadic closure

The clustering coefficient is a measure of the triadic closure of a network of a node neighborhood. The triadic closure is a notion coming from social network analysis, often summarized by the aphorism *The friends of my friends are my friends*.

---

$C_u$ - **Node clustering coefficient:** density of the subgraph induced by the neighborhood of $u$, $C_u = d(H(N_u))$. Also interpreted as the fraction of all possible triangles in $N_u$ that exist, $\frac{\delta_u}{\delta_u^{\max}}$
$\langle C \rangle$ - **Average clustering coefficient:** Average clustering coefficient of all nodes in the graph, $\bar{C} = \frac{1}{N}\sum_{u \in V} C_u$.

> Be careful when interpreting this value, since all nodes contributes equally, irrespectively of their degree, and that low degree nodes tend to be much more frequent than hubs, and their $C$ value is very sensitive, i.e., for a node $u$ of degree 2, $C_u \in 0, 1$, while nodes of higher degrees tend to have more contrasted scores.

$C^g$ - **Global clustering coefficient:** Fraction of all possible triangles in the graph that do exist, $C^g = \frac{3\Delta}{\Delta^{\max}}$

## Small World Network

A network is said to have the **small world** property when it has some structural properties. The notion is not quantitatively defined, but two properties are required:

- Average distance must be short, i.e., $\langle \ell \rangle \approx \log(N)$
- Clustering coefficient must be high, i.e., much larger than in a random network , e.g., $C^g \gg d$, with $d$ the network density

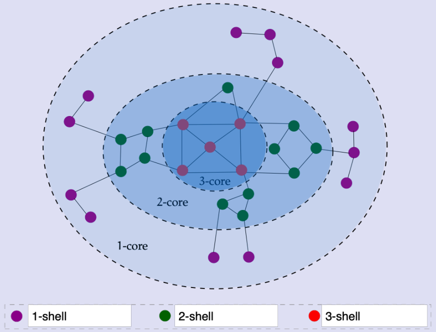
This property is considered characteristic of *real* networks, as opposition to random networks. It is believed to be associated to particular properties (robustness to failures, efficient information flow, etc.), and to be the consequence of emergent mechanisms typical of *complex systems*.

## Cores and Shells

Many real networks are known to have a **core-periphery** structure, i.e., there is a densely connected core at its center and a more peripheral zone in which nodes are loosely connected between them and to the core.

---

**k-core:** The k-core (core of order $k$) of $G(V, E)$ is the largest subgraph $H(C)$ such as all nodes have at least a degree $k$, i.e., $\forall u \in C, k_u^H \leq k$, with$k_u^H$ the degree of node $u$ in subgraph $H$.
**coreness:** A vertex $u$ has coreness $k$ if it belongs to the $k$-core but not to the $k+1$-core.
**c-shell:** all vertices whose coreness is exactly $c$.




## Vocabulary

**Singleton**: node with a degree $k = 0$
**Hub**: node $u$ with $k_u \gg \langle k \rangle$

---

**Bridge**: Edge which, when removed, split a connected component in two.
**Self-loop**: Edge which connects a node to itself.

---

**Complete network**: $L = L_{max}$
**Sparse network**: $d \ll 1$, $L \ll L_{max}$
**Connected Graph**: Graph composed of a single connected component