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.

Directed graph: A weight is associated to every edge.

Network - Graph notation

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

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

Directed graph: Edges have a direction: \((u, v) \in V \) does not imply \((v, u) \in V \)

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, multipletable hypergraphs, etc.) are introduced in sheet ??

Counting nodes and edges

- \( N \): number of nodes \( |V| \)
- \( L \): number of edges \( |E| \)

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 \( |N_u| \).
- \( k^+ \): In-degree of \( u \), number of incoming edges \( |N^+_u| \).
- \( k^- \): Out-degree of \( u \), number of outgoing edges \( |N^-_u| \).
- \( w_{uv} \): Weight of edge \((u, v)\).
- \( s_u \): Strength of \( u \), sum of weights of adjacent edges, \( s_u = \sum_{v \in N_u} w_{uv} \).

Network descriptors - Nodes/Edges

- \( \bar{k} \): Average degree: Real networks are sparse, i.e., typically \( \bar{k} \ll n \). Increases slowly with network size, e.g., \( d \sim \log(m) \).
- \( \ell(G) \): Density: Fraction of pairs of nodes connected by an edge in \( G \).

\( \ell(G) = \frac{2m}{n(n - 1)} \)

Subgraphs

- \( H(W) \): subset of nodes \( W \) of a graph \( G = (V, E) \) and edges connecting them in \( G \), i.e., subgraph \( H(W) = (W, E') \), \( W \subseteq V \), \((u, v) \in E' \iff (u, v) \in E \)
- 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: an undirected subgraph in which any two vertices are connected to each other by paths
- Weakly Connected component: a directed subgraph in which any two vertices are connected to each other by paths if we disregard directions

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).

Network descriptors - Paths

- \( \ell_{\text{max}} \): Diameter: maximum distance between any pair of nodes.
- \( \ell(G) \): Average distance.

\( \ell(G) = \frac{1}{n(n - 1)} \sum d_{ij} \)

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.
- **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.
- **\( \ell_{uv} \): Distance**: The distance between nodes \( u, v \) is the length of the shortest path
Triangles counting

\[ \Delta = \frac{1}{3} \sum_{u \in V} \delta_u. \]

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

\[ \delta_u = \text{Triad potential of } u : \text{maximum number of triangles that could exist around node } u \text{, given its degree}. \]

\[ \Delta_u = \text{Number of triangles containing node } u \] in the graph, total number of triangles in the graph.

\[ \Delta = \frac{1}{3} \sum_{u \in V} \Delta_u. \]

\[ \Delta = \frac{1}{3} \sum_{u \in V} \sum_{v \in V} \sum_{w \in V} \delta_{uvw}. \]

Clustering Coefficients - 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.

\[ C_u = \text{Node clustering coefficient:} \text{density of the subgraph induced by the neighborhood of } u, C_u = \delta(H(N_u)). \text{Also interpreted as the fraction of all possible triangles in } N_u \text{ that exist}. \]

\[ C = \frac{1}{N} \sum_{u \in V} C_u. \]

\[ (C) = \text{Average clustering coefficient:} \text{Average clustering coefficient of all nodes in the graph}, C = \frac{1}{N} \sum_{u \in V} C_u. \]

Be careful when interpreting this value, since all nodes contribute 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.

\[ C^T = \text{Global clustering coefficient:} \text{Fraction of all possible triangles in the graph that do exist}, C^T = \frac{1}{\Delta}. \]

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., \( \langle l \rangle \approx \log(N) \)
- Clustering coefficient must be high, i.e., much larger than in a random network, e.g., \( C^T \gg d \), 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.

\[ k \text{-core: } \text{The } k \text{-core (core of order } k \text{) of } G(V,E) \text{ is the largest subgraph } H(C) \text{ such as all nodes have at least a degree } k, \text{ i.e., } u \in C, k_u^H \leq k, \text{ with } k_u^H \text{ the degree of node } u \text{ in subgraph } H. \]

\[ \text{coreness: } \text{A vertex } u \text{ has coreness } k \text{ if it belongs to the } k \text{-core but not to the } k+1 \text{-core.} \]

\[ c \text{-shell: } \text{all vertices whose coreness is exactly } c. \]

Vocabulary

Singleton: node with a degree \( k = 0 \)

Hub: node \( u \) with \( k_u \gg (k) \)

Bridge: Edge which, when removed, split a connected component in two.

Self-loop: Edge which connects a node to itself.

Complete network: \( L = L_{\text{max}} \)

Sparse network: \( d \ll 1, L \ll L_{\text{max}} \)

Connected Graph: Graph composed of a single connected component