Network Science Cheatsheet

Network Science - Introduction

Networks: Graph notation

Graph notation: \( G = (V, E) \)
- \( V \): set of vertices/nodes.
- \( E \): set of edges/links.
- \( (u, v) \in E \) represents an edge.

Node-Edge description

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

Network descriptors - Paths

- \( \ell_{max} \): Diameter: maximum distance between any pair of nodes.
- \( \langle \ell \rangle \): Average distance:
\[
\langle \ell \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}
\]

Network descriptors - Nodes/Edges

- \( (k) \): Average degree: Real networks are sparse, i.e., typically \( (k) \ll n \). Increases slowly with network size, e.g., \( d \sim \log(n) \).
\[
(k) = 2m/n
\]
- \( d/d(G) \): Density: Fraction of pairs of nodes connected by an edge in \( G \).
\[
d = L/L_{max}
\]

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).

Counting nodes and edges

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

### Examples of networks

- **Simple graph**: Edges can only exist or not exist between each pair of node, and there are no self-loops, i.e., an edge connecting a node to itself.
- **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 later.
Subgraphs

- **Connected component**: subset of nodes $W$ of a graph $G = (V, E)$ and edges connecting them in $G$, i.e., subgraph $H(W) = (W, E')$, $W \subset V, (u, v) \in E' \iff u, v \in W \land (u, v) \in E$.
- **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.

**Weakly Connected component**: In directed networks, a subgraph in which any two vertices are connected to each other by paths.

**Strongly Connected component**: In directed networks, a subgraph in which any two vertices are connected to each other by additional vertices in the supergraph.

- **Bridge**: Edge which, when removed, split a connected component in two.
- **Self-loop**: Edge which connects a node to itself.
- **Stub**: A stub is an half edge, i.e., edge $(u, v)$ has a stub connected to $u$ and another connected to $v$.

- **Singleton**: node with a degree $k = 0$.
- **Hub**: node $u$ with $k_u > k$.

Triangles counting

- **$\delta_u$**: Triads of $u$; number of triangles containing node $u$.
- **$\Delta$**: Number of triangles in the graph.

**Triangles counting**

$$\delta_u = \frac{1}{2} \sum_{v \in V} \delta_{uv}$$

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

$$\Delta = \frac{1}{2} \sum_{u \in V} \delta_u$$

$\Delta$ - Number of triangles in the graph total number of triangles in the graph.

**Triangles counting**

- **$\delta_{\max}$**: maximum number of triangles that could exist around node $u$, given its degree: $\delta_{\max} = \tau(u) = \binom{k}{3}$

**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 $u$, $C_u = \frac{d}{d_u}$.
- **$C$**: Average clustering coefficient: average clustering coefficient of all nodes in the graph, $C = \frac{1}{|V|} \sum_{u \in V} C_u$.

**Clustering Coefficents - Triadic closure**

- 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 $C_i$ 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, $C^g = \frac{2\Delta}{\max} = \Delta$.

Small World Network

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

- Average distance must be short, i.e., $(\ell) \approx \log(N)$.
- 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.

- Watts and Strogatz [1998]

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 (order $k$) of $G(V, E)$ is the largest subgraph $H(C)$ such as all nodes have at least a degree $k$, i.e., $\forall u \in C, k^H_u \geq k$.

**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 $c$.

References