

1

Basics of Network Science

1.1 Basics definition

Lecture 6.
Thursday 15th
October, 2020.
Compiled: Tuesday
27th October, 2020.

The most important thing is the concept of **network** (graph) $G(V, E)$ which is just an object composed by a set of nodes (vertices) V and a set of links (edges) E :

- the **nodes** represents the *entities* $V = [\dots, i, j, k, \dots]$ of the relationships, the entries, the people in a social network and so on. The number of nodes is $N = |V|$;
- the **links** represents the relationship between entities $E = [\dots, (i, j), (i, k), \dots]$. The number of links is $L = |E|$.

Links can be of different types and so networks: the basics distinction is between **undirected** and **directed** links. Another important distinction is between **unweighted** and **weighted** links.

The **network density** (connectance) is the fraction of links over all the possible pairs:

$$d = \frac{L}{N(N-1)} \quad (1.1)$$

Real networks usually have a very low density, so are sparse systems ($L \ll N^2$).

Another things that we will use a lot is the way of representing a graph in a mathematical sense. We will use the **adjacency matrix** A of the network, where:

- $a_{ij} = 1$, if a link between nodes i and j exists;
- $a_{ij} = 0$ otherwise.

We can exploit many useful mathematical properties such as the spectrum of the matrix. Moreover, the matrix is symmetrical for undirected/unweighted graphs, i.e. $a_{ij} = a_{ji}$.

Since real networks are usually sparse, the adjacency matrix is inefficient for storing graphs in a computer, is better to use adjacency lists, etc.

Two nodes that share a link are defined as connected, adjacent, neighbors. In particular, the **neighborhood** of node i is the set of nodes connected to i . The number of neighborhood k_i of each node i is what is called the **degree** of the node. This is the basic measure that we are gonna use a lot. Once you define the degree, the next step is defining what is the average of the degree over the entire network:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i, \quad \text{or} \quad \langle k \rangle = \frac{2L}{N} = d(N-1) \quad (1.2)$$

The next definition is the concept of **path**, which is a sequence of links which permit to go from node i to node j following links. Another important part is what is called the **shortest path** between i and j . This gives us the idea of how big the network is. In particular, the distance l_{ij} represents the length of the shortest path between i and j . There could be multiple shortest paths between i and j .

The network is **connected** if every possible couple of nodes is reachable through a path. Otherwise, each connected part is defined as a connected **component**.

The shortest path of maximum length in the network is defined as **diameter**:

$$l_{max} = \max_{ij} l_{ij}$$

Another measure that is quite important is the **average (shortest) path length**:

$$\langle I \rangle = \frac{\sum_{ij} l_{ij}}{N(N-1)}$$

Now, let us see some example as “The Oracle of Bacon”, the “Erdos Number” and so on. The question which arises is: why such short distances in such large networks? In particular, we note that in the last example of “Six degrees of separation” real networks are smaller (shorter) than one would expect. This is what is called the small world phenomena. If you study the average path length which scales linearly as the scales of the network it scales as the logarithm of the network or in some case the logarithm of the logarithm of the network. This is huge important in the spreading of diseases.