# Lecture 1: Network basics

EDUARDO LÓPEZ

We start our lectures by specifying the basic concepts and nomenclature of networks. It is useful to do this with the aid of a concrete network in mind. In Fig. 1 we display the network of testimonies in the Watergate conspiracy.

**Definition 1.** *A* **network or graph** $G$ *is an object constituted by two sets:*

1. *a node set* $V(G)$ *which is a collection of elements with arbitrary labels, and*

2. *a link set* $E(G)$ *which is a collection of (possibly ordered) pairs of elements from* $V(G)$.

In terms of nomenclature, it is common to use a capital letter such as $G$ to represent a network, but other choices might occur. The nomenclature for the node set $V(G)$ comes from mathematics where nodes are normally called vertices. For the link set, $E(G)$ is typically used, where the name related to the mathematical designation of links as edges.

In the social sciences, the names of nodes and links are typically **actors** and **ties**, respectively.

**Example 1.** *Figure 1 is the visual representation of the directed network* $W$ *of the Watergate testimonies. The node set is given by*

$$\begin{aligned} V(W) &= \{\text{Baldwin}, \text{Barker}, \text{Chapin}, \text{Colson}, \text{Dean}, \text{Ehrlichman}, \text{Gray} \\ &\quad \text{Haldeman}, \text{Hunt(H)}, \text{Kalmbach}, \text{Krogh}, \text{LaRue}, \text{Liddy}, \\ &\quad \text{Magruder}, \text{Martinez}, \text{McCord}, \text{Mitchell}, \text{Nixon}, \text{O'Brien} \\ &\quad \text{Parkinson}, \text{Porter}, \text{Segretti}, \text{Strachan}, \text{Sturgis}\}. \end{aligned}$$

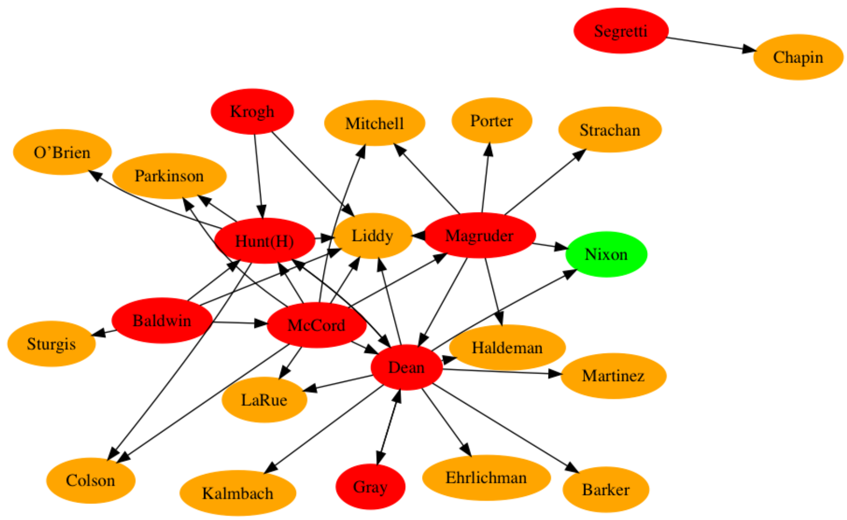


Figure 1: Directed network of the Watergate conspiracy testimonies. The direction of the link indicates who testified against whom. For instance, Segretti testified against Chapin, which is signified by an arrow from Segretti to Chapin. Nodes in red represent individuals that testified against other individuals (represented in orange).

*Similarly, the link set $E(W)$ is given by*

$$\begin{aligned} E(W) \quad = \quad & \{(\text{Baldwin}, \text{Hunt(H)}), (\text{Baldwin}, \text{Liddy}), (\text{Baldwin}, \text{McCord}) \\ & (\text{Baldwin}, \text{Sturgis}), (\text{Dean}, \text{Barker}), (\text{Dean}, \text{Ehrlichman}) \\ & (\text{Dean}, \text{Gray}), (\text{Dean}, \text{Haldeman}), (\text{Dean}, \text{Hunt(H)}) \\ & (\text{Dean}, \text{Kalmbach}), (\text{Dean}, \text{LaRue}), (\text{Dean}, \text{Liddy}) \\ & (\text{Dean}, \text{Martinez}), (\text{Dean}, \text{Nixon}), (\text{Gray}, \text{Dean}), (\text{Hunt(H)}, \text{Colson}) \\ & (\text{Hunt(H)}, \text{Dean}), (\text{Hunt(H)}, \text{Liddy}), (\text{Hunt(H)}, \text{OBrien}), (\text{Hunt(H)}, \text{Parkinson}) \\ & (\text{Krogh}, \text{Hunt(H)}), (\text{Krogh}, \text{Liddy}), (\text{Magruder}, \text{Dean}) \\ & (\text{Magruder}, \text{Haldeman}), (\text{Magruder}, \text{Liddy}), (\text{Magruder}, \text{Mitchell}) \\ & (\text{Magruder}, \text{Nixon}), (\text{Magruder}, \text{Porter}), (\text{Magruder}, \text{Strachan}) \\ & (\text{McCord}, \text{Colson}), (\text{McCord}, \text{Dean}), (\text{McCord}, \text{Hunt(H)}) \\ & (\text{McCord}, \text{LaRue}), (\text{McCord}, \text{Liddy}), (\text{McCord}, \text{Magruder}) \\ & (\text{McCord}, \text{Mitchell}), (\text{McCord}, \text{Parkinson}), (\text{Segretti}, \text{Chapin})\} \end{aligned}$$

The number of nodes of a network is commonly indicated by the symbol $n(G)$ or simply $n$ when the network we are talking about is clear. Similarly,

$m(G)$ or $m$ represent the number of links of a network. Using the concept of cardinality of a set, which is simply the number of elements of a set,

$$\begin{aligned} n &= |V(G)| \\ m &= |E(G)| \end{aligned} \tag{1}$$

where the vertical bars represent cardinality.

Almost all the information related to a network can be captured in a matrix called the **adjacency matrix** of the network. The matrix is the organized collection of a set of numbers that we call *link indicators* $a_{i,j}$, or simply $a_{ij}$ when it is clear that $i$ and $j$ are two distinct nodes. To be precise

**Definition 2.** *A* **link indicator** *is a function $a_{ij}$ of the network $G$ that takes on the value 1 if $(i, j)$ is a link in $G$, i.e., if $(i, j)$ belongs to $V(G)$, and it takes the value 0 if $(i, j)$ is not a link of $G$.*

Note that the function is defined for every pair of nodes $i, j$ of the network, not just for the nodes present (or absent).

Now, when the link indicators are taken together they can be organized in the form of a matrix. It is easiest to imagine this in the context of nodes labelled in a simple sequential way, such as $1, 2, 3, \ldots, n$. Then, to know all the connections that node 1 has, we just need to look at the list of link indicators

$$a_{1,1}, a_{1,2}, a_{1,3}, \ldots, a_{1,n}, \tag{2}$$

where the first indicator checks the connection of node 1 with node 1, then the second indicator checks the connection between node 1 and node 2, etc., all the way to checking the connection of node 1 to node $n$. We have to look at the entire list to be sure we do not miss links. Note that there is a link indicator $a_{1,1}$, which is the indicator for a so-called self-loop, a connection of a node with itself. Although possible in many contexts (imagine a network of emails, and thus somebody emailing themselves for a reminder), we do not use them here for social networks.

In a similar way to node 1, we can see all the connections of any generic node $i$ by looking at

$$a_{i,1}, a_{i,2}, a_{i,3}, \ldots, a_{i,n}. \tag{3}$$

If we collect all such lines for all nodes in a network and organize them as follows

$$\begin{array}{ccccc} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{array} \tag{4}$$

we have pretty much a matrix which only needs the large parenthesis. Therefore

**Definition 3.** *The* **adjacency matrix A of network** $G$ *is the matrix of link indicators*

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{pmatrix} \tag{5}$$

*for the network.*

One piece of information that the adjacency network does not automatically carry with it are the names of the nodes. This is the one piece of information that needs to be tracked independently. Computational tools such as `networkx` track things like that efficiently.

The number of connections a node has is a very important quantity in networks. It has major effects in how influential a node is over the entire network, and usually reflects other underlying features of the node. This quantity is called the *degree of a node*, formally

**Definition 4.** *The* **degree of a node** $i$ *in network* $G$ *is the number of connections that visit* $i$*.*

There are several slight variant of this definition depending on whether the network is *directed* (as it is in the Watergate testimony network), or not directed. Thus, in order to finish defining degree, let us also talk about directed and undirected networks.

When the links of a network are supposed to have direction such as, for instance, in a network describing the hierarchy of authority in a company, the order in which the nodes appear in a link is important. Thus, if $i$ testified against $j$, we write the link as $(i, j)$ and its associated link indicator is $a_{ij}$; the opposite, i.e., that $j$ testified against $i$ is labelled by $(j, i)$ and its link indicator. Often, links that are directed are called arcs.

When a links of a network are not supposed to have direction, such as when two people have been in closed proximity to one another, and that is the only thing we are tracking, then $(i, j)$ and $(j, i)$ are assume to be the same, and their indicators $a_{ij} = a_{ji}$ as well.

For directed networks, degree can be in-degree and out-degree. The in-degree can be found from the link indicators

$$a_{1,i}, a_{2,i}, a_{3,i}, \ldots, a_{n,i}, \tag{6}$$

that is, the connections that start in some other node and reach node $i$. On the other hand, the out-degree can be found by the indicators in Eq. 3. These comments then lead to

**Theorem 1.** *For a directed network $G$ the in-degree of a node $i$ is given by*

$$k^{(in)} = \sum_{j=1}^{n} a_{ji}, \tag{7}$$

*and the out degree of $i$ by*

$$k^{(out)} = \sum_{j=1}^{n} a_{ij}. \tag{8}$$

*For an undirected network, the degree of node $i$ is given by*

$$k = \sum_{j=1}^{n} a_{ji}. \tag{9}$$

Having defined directedness in networks, we can also mention an important property of adjacency matrices for undirected networks: they are symmetric about the diagonal, and thus

**Theorem 2.** *For an undirected network $G$, the adjacency matrix is symmetric about the diagonal.*

Up to this point, we have defined all the links to be equal: a pair of nodes is connected or not. But what about links that represent relationships with different levels of emotional investment, for example? In those cases, it is possible to define *weighted networks*, for which a link has an additional feature beyond simply being present or absent. Thus, we can have an expanded definition for links and the associated link indicator function that reads as follows

**Definition 5.** *A* **weighted link** $(i, j; w)$ *is a link characterized between nodes $i$ and $j$ that also possesses a* **weight feature** $w$ *which provides information about the strength or quality of the link. The link indicator function in weighted networks changes to*

$$a_{ij} = \begin{cases} w & \text{if } (i,j) \text{ is present with weight } w \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In some networks such as infrastructural ones, weights tend to have very straightforwards meanings such as speed limits, capacity to carry traffic, etc. In social networks, they may also have such meanings in quantities such as numbers of messages, time of interaction, etc. In some cases, the weight comes from a scale from a survey ("on a scale of 1-10 how much do you depend on the other person?"). It may be that in some cases, the importance of the scale is not in the number itself for the weight, but in the relative values of them, i.e., is the weight on $(i, j)$ smaller or greater than on $(h, g)$, for example. Finally, we should say that it is typical to write weights with the subindex for the link, so $w_{ij}$ is the weight for the link $(i, j)$.

Some clarifications are in order:

1. The adjacency matrix can also be updated for weighted networks by organizing the link indicators in the same way as for non-weighted matrices.

2. In contrast to weighted networks, non-weighted networks are usually called binary networks, because the links are *either* present or absent.

3. Binary networks are a special case of weighted networks were all links have weight 1.

4. In fact, undirected networks, in a similar way, are also special cases of directed networks where one imposes the condition $a_{ji} = a_{ij}$. This can also be extended to weighted networks where the link indicators have the weights specified.

Another concept that is important to mention is link lists. A link list is an informal, applied way to refer to the link set $E(G)$ of the network $G$. What I mean by informal is that, usually, a link list is found in the context of the data that reflects the network under consideration. For instance, if we open the file in which the Watergate data was stored, we find the text file containing

```
Baldwin Hunt
Baldwin Liddy
Baldwin McCord
Baldwin Sturgis
Dean  Barker
Dean  Ehrlichman
Dean  Gray
Dean  Haldeman
```

```
Dean  Hunt
Dean  Kalmbach
Dean  LaRue
Dean  Liddy
Dean  Martinez
Dean  Nixon
Gray  Dean
Hunt  Colson
Hunt  Dean
Hunt  Liddy
Hunt  OBrien
Hunt  Parkinson
Krogh  Hunt
Krogh  Liddy
Magruder  Dean
Magruder  Haldeman
Magruder  Liddy
Magruder  Mitchell
Magruder  Nixon
Magruder  Porter
Magruder  Strachan
McCord  Colson
McCord  Dean
McCord  Hunt
McCord  LaRue
McCord  Liddy
McCord  Magruder
McCord  Mitchell
McCord  Parkinson
Segretti  Chapin
```

To finish this lecture, it is worth mentioning that, as stated at the beginning, nodes and links actually have numerous names depending on the discipline applying network concepts. Table 1 In these lectures

| Discipline | node | link |
|---|---|---|
| Generic (Network Science) | node | link |
| Mathematics | vertex | edge |
| Social Science | actor | tie |
| Physics | node | bond |

Table 1: Some alternative names for nodes and links