Spatial networks

Carmen Cabrera-Arnau and Elisabetta Pietrostefani

First, what are networks?

Networks are abstract objects used to represent the connections between a set of elements

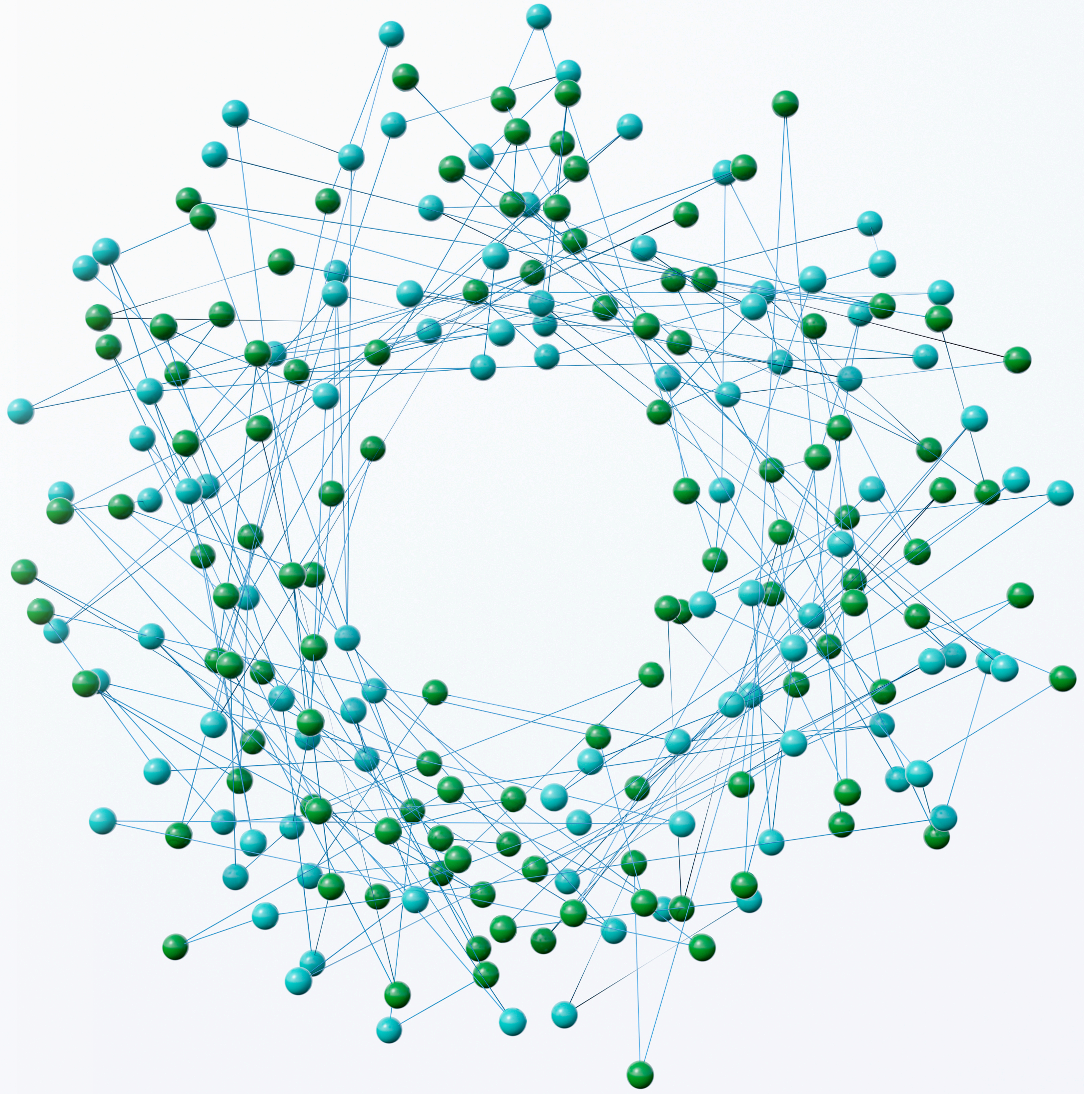
They are in general formed by a set of N **nodes** or vertices and a set of links or **edges** between pairs of nodes

The nodes represent objects, agents, people, locations... you name it! The edges represent some sort of relation between them. For example:

- Nodes are people, edges are friendships
- Nodes are cities, edges are direct railway connections
- Nodes are countries, edges are migratory movements

Depending on the type of nodes and edges, different types of networks can be defined:

- **Simple graph**: there are no nodes with self-links
- **Undirected graph**: all edges are symmetric
- **Directed graph**: one that contains non-symmetric edges (there is a notion of direction)
- **Bipartite graph**: one in which nodes can be divided into two nonempty disjoint subsets, and edges can only be present between pairs of nodes where each node belongs to a distinct category



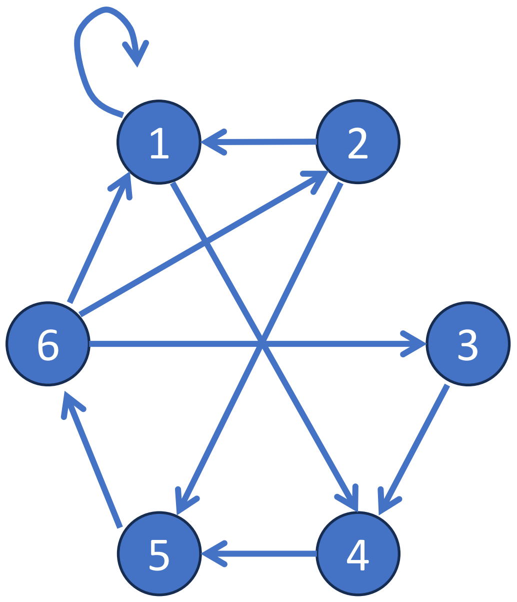
Network representations

- A diagram

- Mathematically, an adjacency matrix
  - In an *N*-node network, there are *N x N* potential links
  - Their presence can be coded by a number which we arrange as the entries of a matrix
  - If there is an edge between node *i* and node *j*, this is normally encoded as a number different from zero in the *ij* element of the adjacency matrix

- Other compact representations

Let's look at an example of adjacency matrix



$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Unweighted, since all the entries are 0 or 1

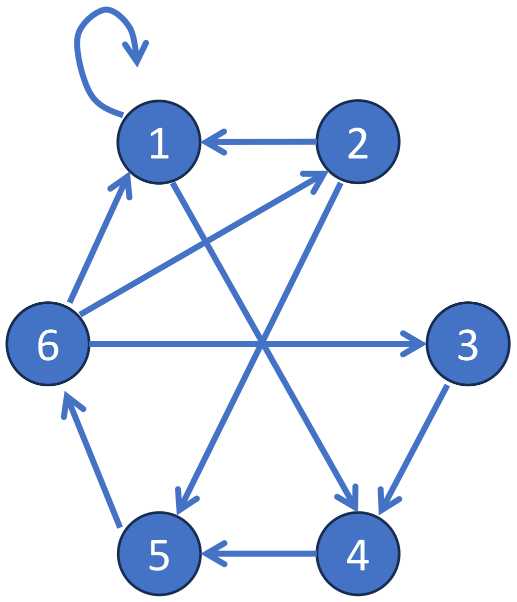
If graph is undirected, the matrix is symmetric

If graph is simple, only zeros in diagonal

If edges have weights, entries are 0 for absent edges or take values $\neq 0$ representing the weights

Disadvantage: as N increases, the matrix scales as $N \times N$

Let's look at an example of adjacency matrix



$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Unweighted, since all the entries are 0 or 1

Alternative representations

As more compact way to represent it is an **adjacency list**

1	1, 4
2	1, 5
3	4
4	5
5	6
6	1, 2, 3

Or simply, **an edge list**

(1, 1), (1, 4),
(2, 1), (2, 5),
(3, 4),
(4, 5),
(5, 6),
(6, 1), (6, 2), (6, 3)

In practice, networks can **conveniently be stored as data frames**

Often network data is in the form of **a data frame of nodes and a data frame of edges**

This allows to store attributes of nodes and edges that can be useful for further analysis, e.g. location of nodes, travel distance through edges, number of people moving from node to node, size of object represented by node, etc.

Data frame of nodes

	Attribute 1: name	Attribute 2: country	Attribute 3: population
Node 1	City 1	Country 1	4500
Node 2	City 2	Country 1	1000
Node 3	City 3	Country 2	2300
…		…	

Data frame of edges

Origin	Destination	Attribute 1: flow	Attribute 2: travel time
Node 1	Node 2	100	45
Node 2	Node 3	60	60
Node 2	Node 5	200	15
…		…	

A network is said to be **spatial** if space is relevant to describe it, i.e. topology or connectivity alone is not enough

In spatial networks, there is a certain notion of “distance”

“Distance” can be:

- Physical distance
- Economical distance
- Social distance
- Administrative distance

Mark Barthélemy (2011). Spatial Networks. *Physics Reports.* 499 (1): 1-101

Hence, the **spatial weights matrix** can be regarded as a type of network, encoding the neighbouring relationships between geographic units

- **Nodes**: the geographic units
- **Edges**: neighbour relationships (according to a certain criterion, e.g. contiguity, distance, etc.)

Conversely, spatial weights matrices can be defined in more general ways according to non-geographical relationships between nodes which have associated locations

An important example of spatial network

A network of flows between locations: **origin-destination matrix**

- Describe flow of individuals, good, information, etc. between locations
- Used since decades by geographers
- Definition
  - Divide the area of interest into zones (cells)
  - Count the number of individuals going from location *i* to location *j*
- Directed
- Weighted
- Strongly depends on zone definition

Types of spatial networks

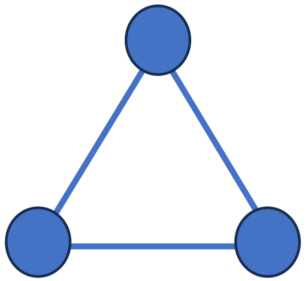
- Transportation networks
  - Airline
  - Bus, subway
  - Cargo ships
- Infrastructure
  - Road and street
  - Power grids, water distribution
- Neural networks
- Protein networks
- Mobility networks
- Social networks

Network metrics

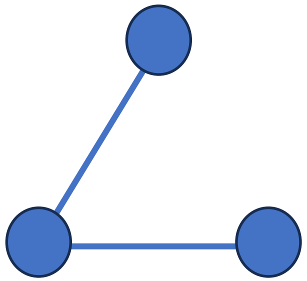
Density: the proportion of existing edges over all possible edges

In an **undirected network** with *N* nodes, the total number of potential edges is $\text{potential edges} = \frac{N \times (N-1)}{2}$

If the number of existing edges is *E,* then the density is of the network is $\frac{existing}{potential} = \frac{2 \times E}{N \times (N-1)}$



Number of nodes = 3
Potential edges = (3x(3-1))/2 = 3
Existing edges = 3
Density = 3/3 = 1



Number of nodes = 3
Potential edges = (3x(3-1))/2 = 3
Existing edges = 2
Density = 2/3 = 0.67

Generally, more dense networks are more robust to shocks (i.e. works in road network less disruptive if there are alternative connections between origin and destination)

Exercise *what would be the formula for the density in a* ***directed network?***

Network metrics

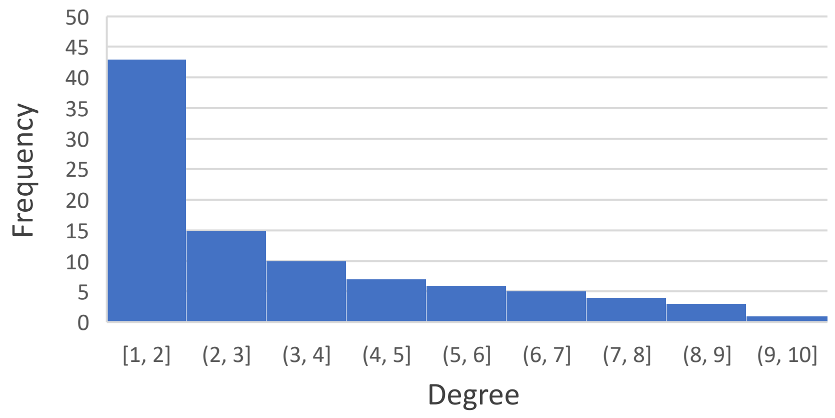
Degree of a node is the number of edges that are connected to it.

In a **directed network**, we can distinguish between in-degree and out-degree

Often, we look at the degree distribution of whole networks (i.e. the number of nodes for each value of the degree)

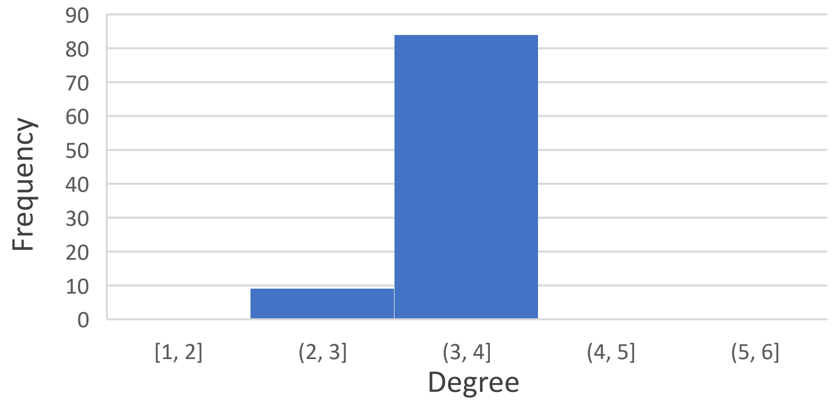
In many real-life networks, the distribution is power-law like (e.g. Pareto distribution):

- In plain English, this means that most nodes have low degree, fewer have high degree and very few have very high degree



For spatial networks, this is not always the case

- For example, a *k* nearest neighbours spatial weights matrix can be regarded as a network, where all nodes have the same degree *k* (except for those in the borders of the region, which may have less than k neighbours)



Network metrics

Shortest path between a pair of nodes is the minimum number of edges that need to be traversed to travel from the origin to the destination node

Diameter of a network is the longest of all shortest paths

Mathematically, a **small-world network** is defined to be a network where the shortest path between two randomly selected nodes grows proportionally to the logarithm of the number of nodes N in the network, that is $\mathrm{L} \propto \log(N)$

- In plain English, a small-world network is one where even when most nodes are not connected with each other, most nodes can be reached from every other by a small number of steps.
- Small-world effect is **common in social networks**, where any two given people are linked by a surprisingly short chain of acquaintances

> ***Six degrees of separation*** *is the idea that two people are six or fewer social connections away from each other. As a result, a chain of friend-of-a-friend statements can be made to connect any two people in a maximum of six steps. It is also known as the* ***six handshakes rule.***

Centrality metrics

Centrality metrics assign scores to nodes (and sometimes also edges) according to their position within a network

These metrics can be used to identify the most influential nodes

Closeness centrality

- A measure of how close a node is to all other nodes in the network
- For a given node, it is computed as the inverse of the average shortest paths between that node and every other node in the network
- Therefore,
  - a node with low value of closeness centrality is close to all the others in the network
  - a node with high value of closeness centrality is far from all the other nodes in the network
- Computing this measure can be costly if the network is large

Betweenness centrality

- A measure of the number of shortest paths going through a node
- Nodes with high values of betweenness centrality indicates that they play a very important role in the overall connectivity of the network
- Can also be computed for edges

Centrality measures are sometimes reported for the whole network as a histogram