# Networks:
# Recap centrality measures

Elsa Arcaute

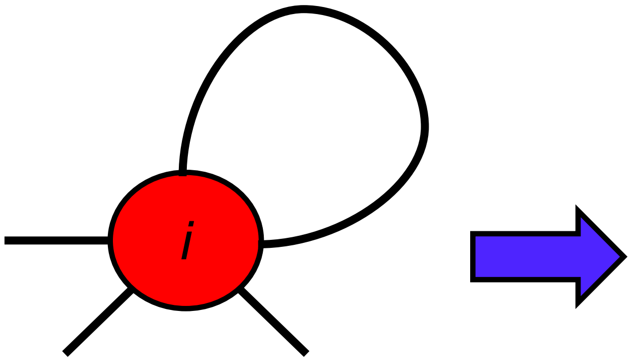
# Outline (Lectures given in 3 parts)

1. Definitions: node, link, degree, etc.
2. Undirected, weighted, directed networks
3. Adjacency matrix, paths and connectivity
4. Centrality measures: closeness, betweenness, etc.
5. Clustering, similarity and modularity
6. Degree distribution
7. Scale-free network
8. Preferential attachment
9. Random graphs
10. Small-world
11. Community detection
12. Spatial networks

**Last lecture**

# Degree of a node: undirected case

The degree of a node is given by the number of links attached to it.
→ The degree of a node *i* is denoted as $k_i$




$k_i=5$

$$\boxed{k_i = \sum_{j=1}^{n} A_{ij}}$$

e.g.

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & \cdots & \cdots & A_{26} \\ A_{31} & A_{32} & \cdots & \cdots & \cdots & \cdots \\ A_{41} & \cdots & \cdots & \cdots & \cdots & \cdots \\ \boxed{A_{51} \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots} \\ A_{61} & \cdots & \cdots & \cdots & \cdots & A_{66} \end{pmatrix} \longrightarrow k_5 = \sum_{j=1}^{6} A_{5j} = A_{51} + \ldots + A_{56}$$

## Mean degree, number of links: undirected case

In an undirected graph, each link has associated two end points.
→ Each link contributes to two nodes' degree
→ Total sum of all degrees = twice the number of links (m)

$$2m = \sum_{i=1}^{n} k_i \quad \Rightarrow \quad m = \frac{1}{2}\sum_{i=1}^{n} k_i = \frac{1}{2}\sum_{ij} A_{ij}$$

Mean degree <*k*> of a node, recall m=number of links and n=number of nodes

$$\boxed{< k > = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n}}$$

# Degree directed case: in/out-degree

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$

Number of links pointing to you

Number links you point to others

**For a directed graph you need to consider these two degrees separately**

## In/out-degree

- → Since the matrix is symmetric for an undirected network, the sum can be over the row or over the column
- → For directed networks care must be taken!

$$k_j^{out} = \sum_{i=1}^{n} A_{ij}$$

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}$$

$$k_i = \sum_{j=1}^{n} A_{ij}$$

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & \cdots & \cdots & A_{26} \\ A_{31} & A_{32} & \cdots & \cdots & \cdots & \cdots \\ A_{41} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{51} & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{61} & \cdots & \cdots & \cdots & \cdots & A_{66} \end{pmatrix}$$

## Mean degree, number of links: directed case

In a directed graph, each link contributes to ONLY ONE node's degree: **in OR out**
→ Total sum of all **in or out degrees** = total number of links

$$m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} = \sum_{ij} A_{ij}$$

Mean in/out-degree of a node: $<k^{in}> = <k^{out}> = <k>$

$$< k^{in} > = \frac{1}{n} \sum_{i=1}^{n} k_i^{in} = \frac{1}{n} \sum_{j=1}^{n} k_j^{out} = < k^{out} >$$

$$\boxed{< k > = \frac{m}{n}}$$

## Centrality Measures

1. **Degree centrality**: recall for directed networks need to compute in- and out-degree. They do not convey the same information!!!!

$$k_i = \sum_{j=1}^{n} A_{ij}$$

If you would like to compare graphs of different sizes, you need to normalise the measure:

$$C_i^d = \frac{k_i}{N-1}$$

## 2. Eigenvector centrality: x

→ It's not about how many connections you have, but *how important* those connections are!

$$x_i = \kappa_1^{-1} \sum_j A_{ij} x_j$$

First eigenvalue of adjacency matrix

✧ Encounter issues with directed networks

→ which direction should we use?

**use nodes pointing at you!**

Note adjacency matrix non-symmetric for directed graphs

# 3. Katz centrality: x

→To avoid null centrality when links point to you: add a bit of centrality "for free" to each node **β**

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

$$\mathbf{x} = \alpha\, \mathbf{A}\mathbf{x} + \beta\, \mathbf{I}$$

↓

$$\mathbf{x} = (\mathbf{I} - \alpha\, \mathbf{A})^{-1}\, \mathbf{1}$$

→ Setting β=1

→ Note that $\alpha \leq 1/k_1$. If this is very close to = then it gives you the eigenvector centrality with non-zero terms for very low centrality.

# 4. Closeness centrality

Let $d_{ij}$ be the geodesic between i and j. The mean geodesic distance is:

$$l_i = \frac{1}{n}\sum_j d_{ij}$$

n, is the total number of nodes.

A person that is very close to most nodes, and has hence low mean geodesic, will be influential: we define closeness centrality as

$$\boxed{C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}}$$

## 5. Betweenness centrality

Let us normalise the measure so that it lies between 0 and 1

$$\boxed{x_i = \frac{1}{n^2} \sum_{st} \frac{n^i_{st}}{g_{st}}}$$

n, is the total number of nodes.

where $g_{st}$ is the total number of geodesics from $s$ to $t$

$$n^i_{st} = \begin{cases} 1 & \text{if vertex } i \text{ lies on geodesic path from } s \text{ to } t \\ 0 & \text{otherwise} \end{cases}$$

→ Flow of information or any other sort of traffic assuming that this takes the shortest path!!!!

→ In a real system might have to modify this betweenness according to the more realistic behaviour of the system, where the other extreme would be a random walk.

→ High betweenness indicates the role of a *broker* in the system

## 6. Delta-centrality (Vito Latora and Massimo Marchiori)

$$C_i^{\Delta} = \frac{(\Delta P)_i}{P} = \frac{P[G] - P[G']}{P[G]}$$

Where P[G] is the performance of a graph G, G' is the new graph after removing *i* and P[G'] is the performance of the new graph G';
$(\Delta P)_i$ is the variation of the performance after deactivation of node *i*

How to choose P? It must satisfy: $(\Delta P)_i \geq 0$

What about efficiency?

$$E = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq 1}^{N} \frac{1}{d_{ij}}$$

N is the total number of nodes.

→ If the efficiency between two nodes *i* and *j* is $1/d_{ij}$, E is the average over all pairs.