Measures and Metrics, Nodes

Measures and Metrics

Given a chosen *metrics* to simplify and represent a studied network, we could count on our innate ability to find patterns from a visualisation of the network to discover some facts of the network by inspecting it.

However, this approach does not scale the larger the network gets.

A better approach is to define *mathematical measures* that capture interesting features of network structure quantitatively, boiling down large volumes of complex structural data into numbers that are an indication of the studied phenomena.

Kinds of Metrics • Binary Scale

The simplest and most popular kind of metrics. Conventionally, 1 indicates the presence of a relationship and 0 indicates its absence.

Being the “ground floor" of the information, it can always be obtained starting from another metric, defining a threshold value (cut-off point) below which all values are reported to 0 and above to 1.

The information that is lost in this way is often compensated by the greater ease of analysis.

Kinds of Metrics • Multi-category Nominal Scales

This metric indicates for each relation the "type" that it assumes, with respect to multiple-choice list (example: lover, friend, colleague, enemy, ...).

The analysis can be carried out at the level of a single type (e.g., networks that have "lover" as a link between the nodes), with effects on the measures (e.g., reduction of density) of which it is important to be aware.

Kinds of Metrics • Ordinal Scales

The simplest ordinal metric refers to a three-value scale, of the type “-1 0 +1”, where:

- - 1 implies the presence of a “negative” relationship (e.g., “aversion of one actor to another”);
- 0 indicates indifference;
- +1 implies the symmetrical situation to the negative one.

Other ordinal measures refer to larger scales, e.g., the Likert one or based on the request to each actor to express the order with which (s)he would like to have relations with the other nodes of the network.

Ordinals can always be brought back to one of the previous scales, losing information.

Kinds of Metrics • Scalar Scales

Scalar metrics are useful when handling values representing either physical quantities - like metres, kilograms, seconds, amperes, moles - or information units and units of account - money, goods, services, assets, labor, income, expenses.

Scalar measures have been developed more recently, through the adaptation of algorithms originally created for the much simpler binary metrics.

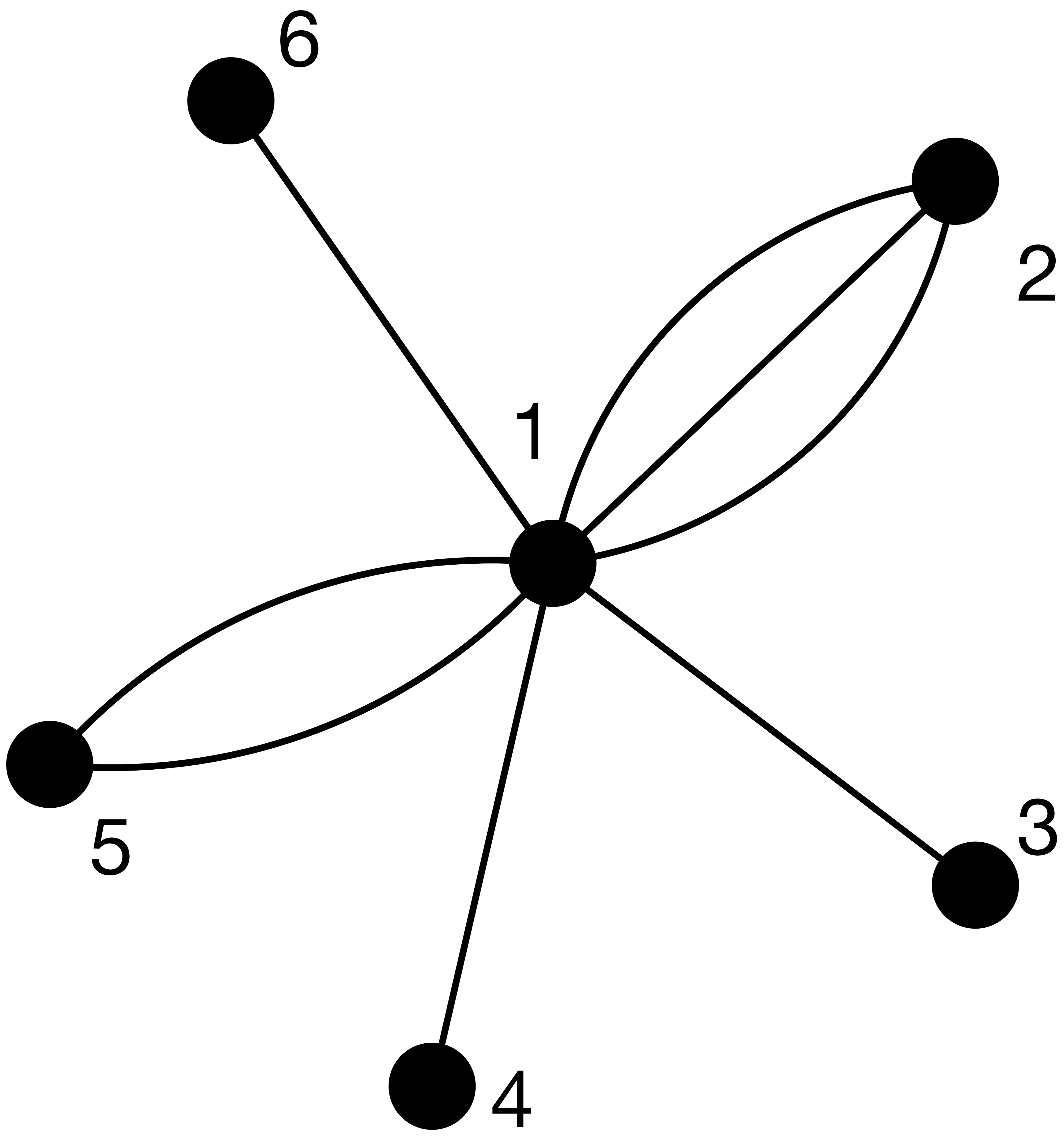
Degree

One of the simplest measures is the degree of a node.

In an undirected network, the degree of a node is the **number of edges connected** to it.

E.g., in a social network of friendships between individuals a person's degree is the number of friends they have.

Despite its simplicity, the degree is one of most useful and most widely used of network concepts and it plays an important role in other measures.



$$deg(1) = 8 \ \cdots \ deg(5) = 2 \ \cdots \ deg(3) = 1$$

$$deg(i) = \sum_{j=1}^{n} A_{ij}$$

Centrality

Centrality measures answer to the question:

"Which are the most important or central nodes in a network?"

Of course, there are many possible definitions of "importance" and there are correspondingly many centrality measures for networks.

Centrality • Degree Centrality

One of the simplest centrality measure for a node in a network is just its **degree**.

In **directed networks**, nodes have both an **in-degree** and an **out-degree**, and both may be useful as measures of centrality in the appropriate circumstances.

Although degree centrality is a simple centrality measure, it can be very illuminating.

For example, in a social network those individuals who have many <u>followers</u> might have more <u>influence</u>, more <u>access to information</u>, or <u>more prestige</u> than those who have fewer.

A non-social network example is the use of citation counts in the evaluation of scientific papers. The number of citations a <u>paper</u> receives from other papers, which is its <u>in-degree</u> in the directed citation network, gives a quantitative measure of how influential the paper is.

Centrality • Eigenvector Centrality

In many circumstances a node's importance in a network is increased by having connections to other nodes that <u>*are themselves important*</u>.

For example, you might have only one friend in the world, but if that friend is the president of the United States then you yourself may be an important person. Thus centrality is not only about how many people you know but also who you know.

Eigenvector centrality is an extension of degree centrality that takes this factor into account. Instead of just awarding one point for every network neighbour a node has, eigenvector centrality awards a number of points ***proportional to the centrality scores of the neighbours***.

Centrality • Eigenvector Centrality

Considering an undirected network of *n* nodes, the eigenvector centrality x_i of node *i* is proportional to the sum of the eigenvectors centralities of *i*'s neighbours.

Mathematically

$$x_i \mathrel{\tilde{=}} \sum_{j \in neighbours(i)} x_j$$

Since it is a sum, a node can achieve a high eigenvector centrality either by having a lot of neighbours with modest centrality or a few neighbours with high centrality (and everything in between) - the intuitive interpretation of this is that nodes can be influential either by reaching a lot of nodes or by reaching just a few, highly-influential nodes.

Centrality • Eigenvector Centrality

Considering an undirected network of *n* nodes, the eigenvector centrality x_i of node *i* is proportional to the sum of the eigenvectors centralities of *i*'s neighbours.

Mathematically

Adjacency Matrix Eigenvector Eigenvector

$$\underbrace{A}_{\text{Adjacency Matrix}}\,\underbrace{\mathbf{x}}_{\text{Eigenvector}} = \underbrace{\kappa}_{\text{Eigenvalue}}\,\underbrace{\mathbf{x}}_{\text{Eigenvector}}$$

Eigenvalue

Since it is a sum, a node can achieve a high eigenvector centrality either by having a lot of neighbours with modest centrality or a few neighbours with high centrality (and everything in between) - the intuitive interpretation of this is that nodes can be influential either by reaching a lot of nodes or by reaching just a few, highly-influential nodes.

Centrality • Eigenvector Centrality

Considering an undirected network of *n* nodes, the eigenvector centrality x_i of node *i* is proportional to the sum of the eigenvectors centralities of *i*'s neighbours.

Mathematically

$$A\mathbf{x} = \kappa\mathbf{x} \quad \Rightarrow \quad x_i = \kappa^{-1} \sum_{j \in neighbours(i)} x_j \quad \Rightarrow \quad x_i = \kappa^{-1} \sum_{j=1}^{n} A_{ij}\, x_j$$

Since it is a sum, a node can achieve a high eigenvector centrality either by having a lot of neighbours with modest centrality or a few neighbours with high centrality (and everything in between) - the intuitive interpretation of this is that nodes can be influential either by reaching a lot of nodes or by reaching just a few, highly-influential nodes.

Centrality • Eigenvector Centrality

Considering an undirected network of *n* nodes, the eigenvector centrality x_i of node *i* is proportional to the sum of the eigenvectors centralities of *i*'s neighbours.

Mathematically

$$\underbrace{A}_{\text{Adjacency Matrix}}\underbrace{\mathbf{x}}_{\text{Eigenvector}} = \underbrace{\kappa}_{\text{Eigenvalue}}\underbrace{\mathbf{x}}_{\text{Eigenvector}}$$

Since it is a sum, a node can achieve a high eigenvector centrality either by having a lot of neighbours with modest centrality or a few neighbours with high centrality (and everything in between) - the intuitive interpretation of this is that nodes can be influential either by reaching a lot of nodes or by reaching just a few, highly-influential nodes.

Centrality • Eigenvector Centrality

How do we choose κ?

Thanks to Perron and Frobenius, we know that

for a square matrix with all elements non-negative (like our adjacency matrix A) there exists a unique largest eigenvalue (κ) and the corresponding eigenvector ($\mathbf{x}$), called leading, can be chosen to have strictly positive components

This means that the eigenvector centrality x_i of node *i* is the *i*-th element of the leading eigenvector of the adjacency matrix and the value of the constant κ is the leading eigenvalue.

Centrality • Eigenvector Centrality

Although we fixed κ, $\mathbf{x}$ remains arbitrary within a multiplicative constant.

Since we are working with eigenvalues/vectors, the **multiplicative constant does not matter** much - we are applying transformations to the values in our adjacency matrix that maintain their proportions.

Practically, this means that, using eigenvector centrality to pick out the most important nodes in a network or to rank nodes from most to least important is **independent from the absolute values of the nodes** and it is only the relative scores of different nodes that matter.

However, if absolute numbers matter (e.g., when comparing different matrices) we can normalise the centralities by, for instance, requiring that they sum to n (which ensures that average centrality stays constant as the network gets larger).

Centrality • Eigenvector Centrality

For the case of **directed networks**, the eigenvector centrality poses some complications due to the **asymmetricity** of adjacency matrices. This translates into two sets of eigenvectors, left and right, and two leading eigenvectors.

Which to choose among the two depends on the reason of the calculation of the centrality measure. The **right eigenvector** measures centrality as **bestowed by others** to the node. The **left eigenvectors** measures centrality as **connections of the node to the others**.

For example, in the Web and in citation networks, a good indication of the importance of a node is how many nodes point to it. However, if we consider transport networks, hubs that connect to a lot of locations tend to be more important.