Network Analysis

AN INTRODUCTION FOR HUMANISTS

Dr Katarzyna Anna Kapitan

13 March 2025

Recap

- Hubs & Centrality
- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Robustness

Centralities

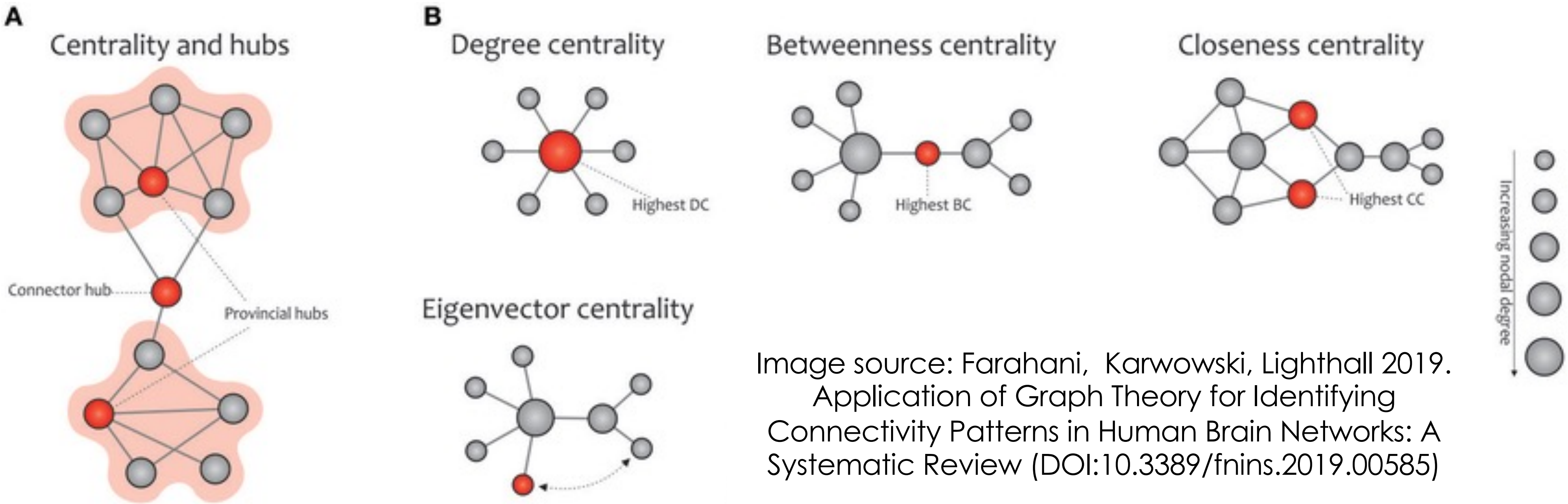


Image source: Farahani, Karwowski, Lighthall 2019. Application of Graph Theory for Identifying Connectivity Patterns in Human Brain Networks: A Systematic Review (DOI:10.3389/fnins.2019.00585)

Centralities

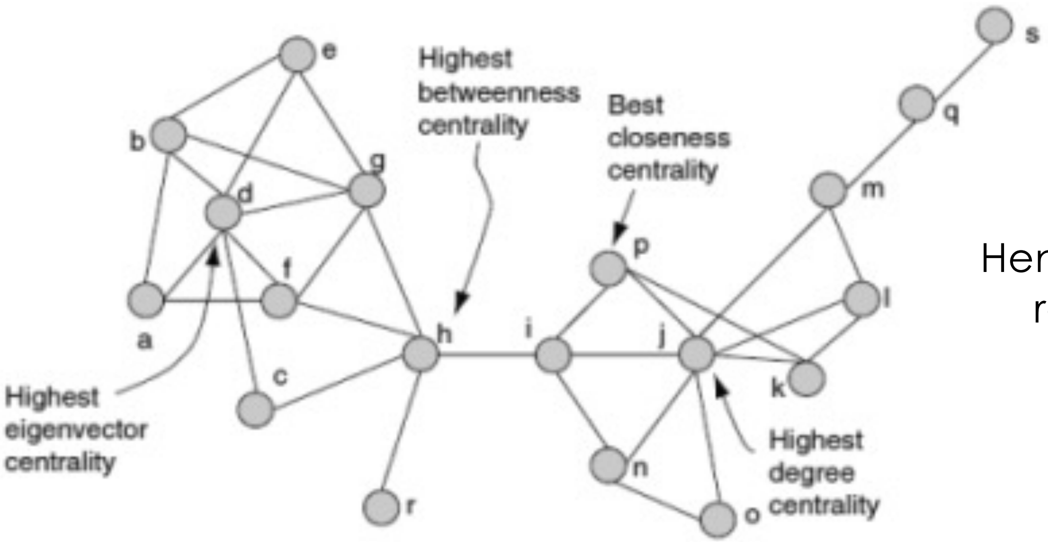


Image Source:
Hendratno & Fitriati 2016. The study of Indonesia's readiness to cope with demographic bonus (DOI:10.22146/jieb.10311)

Measuring Importance and Influence

Degree centrality – assigned a point to a node (vertex) for every neighbour in the network.

‘But not all neighbours are equivalent. In many circumstances a vertex's importance in a network is increased by having connections to other vertices that are *themselves important*.’
(Newman 2010, *Networks: An Introduction*)

In politics, for example, being friends with the president of a country might be more important than being friends with 10 people at the bottom of the administrative pyramid. **How to measure that?**

Eigenvector Centrality

Eigenvector Centrality
(**eigencentrality** or **prestige score**)
is a common measures used to quantify node's importance and influence in the network.

'Instead of awarding vertices just one poin for each neighbour, eigenvector centrality gives each vertex a score proportional to t sum of the scores of its neighbours'
(Newman 2010, *Networks: An Introduction*)

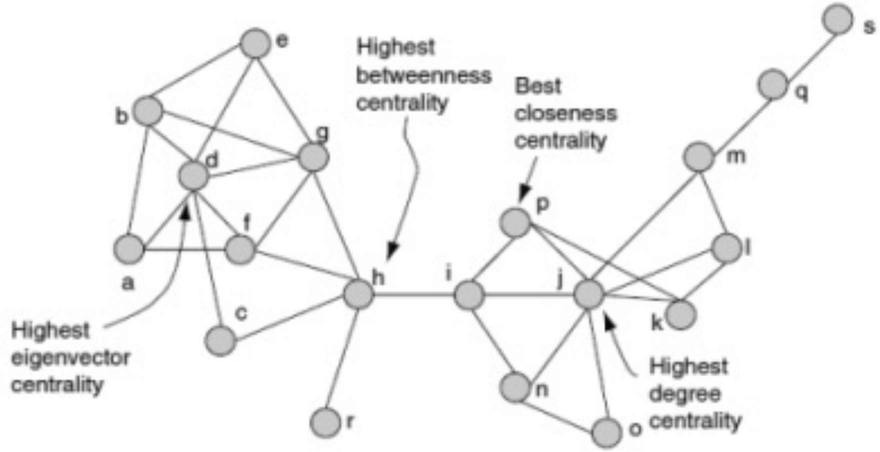


Image Source:
Hendratno & Fitriati 2016. The study of Indonesia's readiness to cope with demographic bonus
(DOI:10.22146/jieb.10311)

Eigenvector Centrality

The eigenvector centrality of node *i* (vertex *i*) is proportional to the sum of the centralities of *i's neighbours*.

‘The eigenvector centrality [has] the nice property that it can be large either because **a vertex has many neighbours** or because it **has important neighbours** (**or both**). An individual in a social network, for instance, can be important, by this measure, because he or she knows lots of people (even though those people may not be important themselves) or knows a few people in high places.’ (Newman 2010, *Networks: An Introduction*)

A high eigenvector score means that a node is connected to many nodes who themselves have high scores.

Eigenvector Centrality

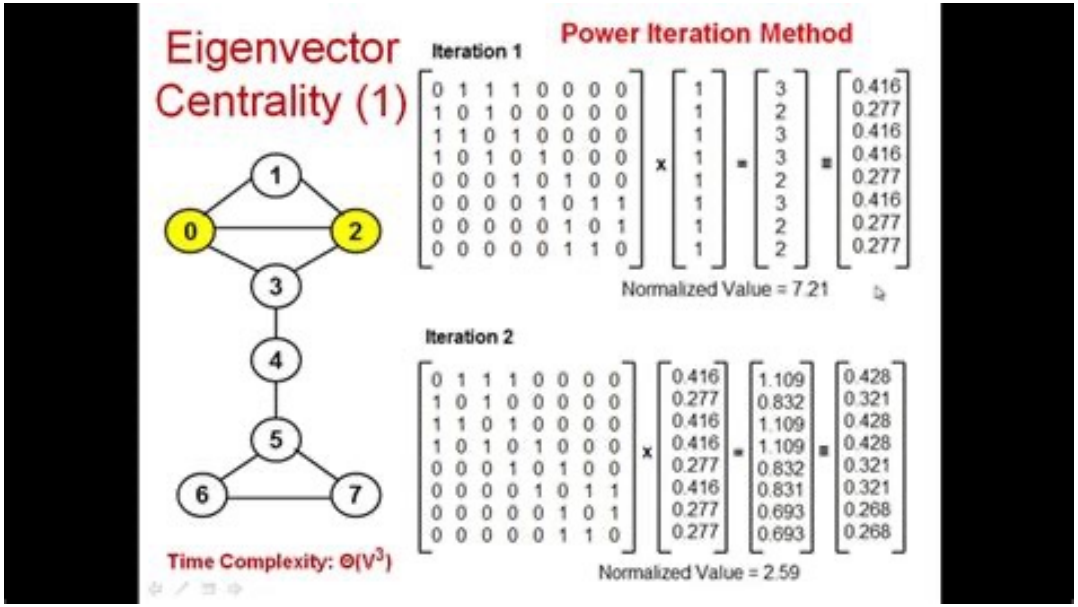
For a given graph let $\boldsymbol{A} = (a_{i,j})$ be the adjacency matrix, where $a_{i,j}=1$ if vertex i is linked to vertex j, and $a_{i,j}=0$ otherwise.
The relative centrality score, x_i, of vertex i can be defined as:

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

Where κ_i are the eigenvalues of $\boldsymbol{A}$, and κ_1 is the largest of them.

(Newman 2010, Eq. 7.6)

Eigenvector Centrality (watch at home)



Natarajan Meghanathan, Eigenvector Centrality,
https://www.youtube.com/watch?v=AjacGCIQ56o

Eigenvector Centrality

In NetworkX, there are functions to compute the eigenvector centrality for the graph by adding the centrality of its predecessors. We will work with them in Lab5.

eigenvector_centrality_numpy(G, weight=None, max_iter=50, tol=0)

eigenvector_centrality(G, max_iter=100, tol=1e-06, nstart=None, weight=None)

Source: https://networkx.org/documentation/stable/reference/algorithms/centrality.html

Eigenvector Centrality

Eigenvector centrality can be calculated for both undirected or directed networks, but it is more commonly used in undirected networks.

In a directed network the adjacency matrix is asymmetric, so it has two sets of eigenvectors, left and right eigenvectors, which allow you to measure how the node is **influenced** by others, and how it **influences** others.

You can think so them as analogous to in-degree and out-degree but more complex, as they capture the global structure of the network taking into account all nodes in your system.

Eigenvector Centrality

‘The **centrality** in **directed networks** is usually bestowed **by other nodes pointing towards you**, rather than by you pointing to others (so incoming links rather than outgoing links).

On the World Wide Web, for instance, the number and importance of web pages that point to your page can give a reasonable indication of how important or useful your page is.

The fact that your page might point to other important pages is irrelevant. Anyone can set up a page that points to a thousand others, but that does not make the page important.’ Source: Newman 2010, *Networks: An Introduction*

Eigenvector Centrality

‘**Vertex A** in [Figure 7.1] is connected to the rest of the network but has **only outgoing edges** and no incoming ones. Such a vertex will always **have centrality zero** because there are no terms in the sum [Eq. 7.6] defining eigenvector. […]

This might not seem to be a problem: perhaps a vertex that no one points to *should* have centrality zero.

But then consider **vertex B**, which has **one ingoing edge**, but that edge originates at vertex A, and hence B also has **centrality zero**, because the one term in its sum [Eq. 7.6] is zero.’ Newman 2010, *Networks: An Introduction*

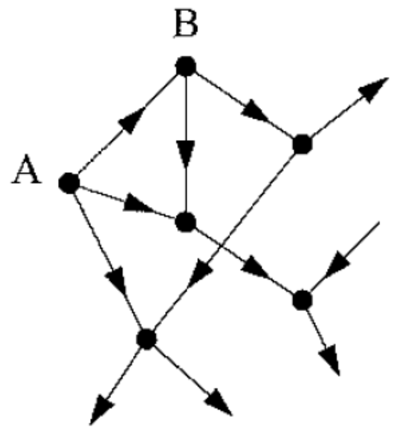


Figure 7.1: A portion of a directed network. Vertex A in this network has only outgoing edges and hence will have eigenvector centrality zero. Vertex B has outgoing edges and one ingoing edge, but the ingoing one originates at A, and hence vertex B will also have centrality zero.

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

(Newman 2010, Eq. 7.6)

Eigenvector Centrality

Taking this argument further, we see that a vertex may be pointed to by others that themselves are pointed to by many more, and so on through many generations, but if the progression ends up at a vertex or vertices that have in-degree zero, it is all for nothing-the final value of the centrality will still be zero.

(Newman 2010, *Networks: An Introduction*)

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

(Newman 2010, Eq. 7.6)

Katz Centrality

Katz Centrality - solution to the simple eigenvector centrality problem. It gives each vertex a small amount of centrality 'for free' (*beta*) regardless of its position in the network (Eq. 7.8)

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

(Newman 2010, Eq. 7.8)

'The first term is the **normal eigenvector centrality** term [compare with Eq. 7.6] in which the centralities of the vertices linking to *i* are summed, and the second term is the "free" part [*beta*], the constant extra term that all vertices receive. By adding this second term, even vertices with zero in-degree still get centrality *beta* and once they have a non-zero centrality, then the vertices they point to derive some advantage from being pointed to. **This means that any vertex that is pointed to by many others will have a high centrality, although those that are pointed to by others with high centrality themselves will still do better**.' (Newman 2010, *Networks: An Introduction*)

Katz Centrality

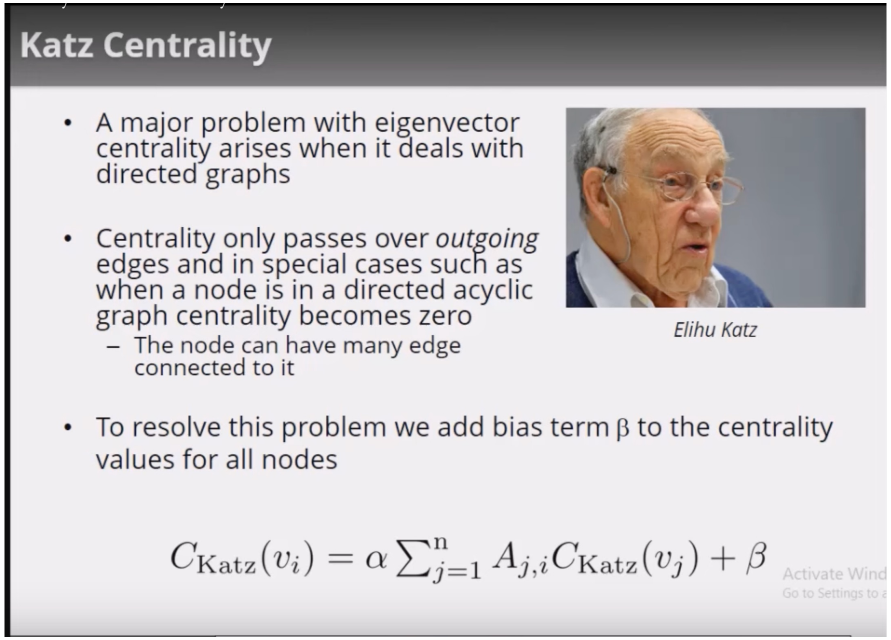
In NetworkX there are two functions to calculate katz centrality and we will test them in Lab5.

katz_centrality_numpy(G, alpha=0.1, beta=1.0, normalized=True, weight=None)

katz_centrality(G, alpha=0.1, beta=1.0, max_iter=1000, tol=1e-06, nstart=None, normalized=True, weight=None)

Source: https://networkx.org/documentation/stable/reference/algorithms/centrality.html

Katz Centrality (Watch at home)



Shreesudha Kembhavi, EigenVector Centrality & Katz Centrality, https://www.youtube.com/watch?v=-LO9NLaccFQ

Katz Centrality

Issue with Katz:

‘If a vertex with high Katz centrality points to many others then those others also get high centrality. A high-centrality vertex pointing to one million others gives all one million of them high centrality. One could argue-and many have-that this is not always appropriate’ (Newman 2010, *Networks: An Introduction*).

Solution:

Make it proportional to out-degree → **PageRank**
‘Then vertices that point to many others pass only a small amount of centrality on to each of those others, even if their own centrality is high’ (Newman 2010, *Networks: An Introduction*)

PageRank

'PageRank works on the Web precisely because having links to your page from important pages elsewhere is a good indication that your page may be important too. But the added ingredient of dividing by the out-degrees of pages insures that pages that simply point to an enormous number of others do not pass much centrality on to any of them'. (Newman 2010, *Networks: An Introduction*)

The hubs do not have a disproportionate influence on the rankings of nodes around them.

PageRank

The formula for PageRank contains one the parameter *alpha,* whose value must be chosen before the algorithm can be used.
Its value should be less than the inverse of the largest eigenvalue [compare Eq. 7.6 and 7.8].

'The *Google* search engine uses a value of ***alpha* = 0.85** in its calculations, although it's not clear that there is any rigorous theory behind this choice. More likely it is just a shrewd guess based on experimentation to find out what works well.' (Newman 2010, *Networks: An Introduction*)

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{\text{out}}} + \beta.$$

(Newman 2010, Eq. 7.15)

PageRank

In NetworkX there is a function pagerank, which we will explore in Lab5.

> **pagerank**(G, alpha=0.85, personalization=None, max_iter=100, tol=1e-06, nstart=None, weight='weight', dangling=None)

Note: When working with PageRanks, remember to remove self-loops from your data & remember that PageRanks across your nodes should equal to 1.

Source:https://networkx.org/documentation/stable/reference/algorithms/link_analysis.html#module-networkx.algorithms.link_analysis.pagerank_alg

Weight

Weight

Many real-world networks (directed and undirected) have link weights:
- Twitter/BlueSky diffusion: number of retweets between two users
- Email: number of messages
- Air transportation: number of passengers

You can think of weights even in networks that we usually represent as unweighted:
- Facebook friendships could be weighted by number of interactions (tags, likes, comments, common friends)
- The movie co-star network could be weighted by number of movies in which two stars have acted together
- Wikipedia links can be weighted by number of clicks

In weighted networks degree, in-degree and out-degree
turn into **strength**, **in-strength**, and **out-strength**

Weights and Influence

Weight is useful when, for example, working with Information Diffusion Networks.
To measure the influence of a an account in SoMe network we can use:

- Number of followers (in-degree in follower network)
- Number of users exposed (out-degree in retweet network)
- Fraction of retweets to followers
- **Number of retweets (out-strength in retweet network)**

Weights and Influence

eigenvector_centrality_numpy(G, **weight=None**, max_iter=50, tol=0)

eigenvector_centrality(G, max_iter=100, tol=1e-06, nstart=None, **weight=None**)

katz_centrality_numpy(G, alpha=0.1, beta=1.0, normalized=True, **weight=None**)

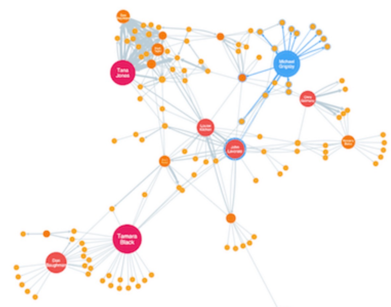
katz_centrality(G, alpha=0.1, beta=1.0, max_iter=1000, tol=1e-06, nstart=None, normalized=True, **weight=None**)

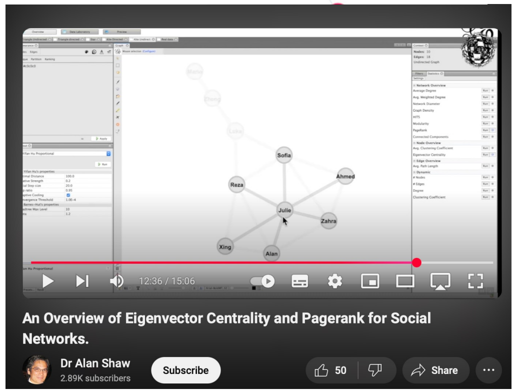
pagerank(G, alpha=0.85, personalization=None, max_iter=100, tol=1e-06, nstart=None, **weight='weight'**, dangling=None)

Still a bit lost?

At home,
- read this short blog post: Andrew Disney, ‘PageRank centrality & EigenCentrality’, 2020, https://cambridge-intelligence.com/eigencentrality-pagerank/

- watch this video demonstrating the difference between Eigenvector anc PageRank using *Gephi:* Dr Alan Shaw, ‘An Overview of Eigenvector Centrality and Pagerank for Social Networks’, 2019, https://www.youtube.com/watch?v=kmQkXIg6Drw





Install *Gephi* before we meet next week!