

# Centrality Measures Based on Matrix Functions

Lembris Laanyuni Njotto

Department of Mathematics and ICT, College of Business Education, Dar Es Salaam, Tanzania

Email: l.njotto@cbe.ac.tz

**How to cite this paper:** Njotto, L.L. (2018) Centrality Measures Based on Matrix Functions. *Open Journal of Discrete Mathematics*, 8, 79-115.

<https://doi.org/10.4236/ojdm.2018.84008>

**Received:** July 12, 2018

**Accepted:** September 23, 2018

**Published:** September 26, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Network is considered naturally as a wide range of different contexts, such as biological systems, social relationships as well as various technological scenarios. Investigation of the dynamic phenomena taking place in the network, determination of the structure of the network and community and description of the interactions between various elements of the network are the key issues in network analysis. One of the huge network structure challenges is the identification of the node(s) with an outstanding structural position within the network. The popular method for doing this is to calculate a measure of centrality. We examine node centrality measures such as degree, closeness, eigenvector, Katz and subgraph centrality for undirected networks. We show how the Katz centrality can be turned into degree and eigenvector centrality by considering limiting cases. Some existing centrality measures are linked to matrix functions. We extend this idea and examine the centrality measures based on general matrix functions and in particular, the logarithmic, cosine, sine, and hyperbolic functions. We also explore the concept of generalised Katz centrality. Various experiments are conducted for different networks generated by using random graph models. The results show that the logarithmic function in particular has potential as a centrality measure. Similar results were obtained for real-world networks.

## Keywords

Graph, Centrality Measures, Matrix Functions, Kendall Correlation Coefficient, Random Graph Models

## 1. Introduction

Since its introduction by Euler in eighteen century, graph theory has proven its important applications in many different scientific fields. Graphs and linear algebra have been used to model social interactions. Recently network models are now commonplace not only in the hard sciences but also in various technologi-

cal, social and biological scenarios. Networks are used to model a variety of highly interconnected systems, both in nature and man-made world of technology. These networks include protein-protein interaction networks, social networks, food webs, scientific collaboration networks, metabolic networks, lexical or semantic networks, neural networks, the World Wide Web and others. The use of network analysis is in various situations: from determining network structure and communities, describing the interactions between various elements of the network and investigating the dynamics phenomena taking place in the network [1].

One of the ground laying questions analysis of network is how to determine the “most important” nodes in a given network. Many centrality measures have been proposed, starting with the simplest of all, node degree centrality. This measure has being considered too “local”, as it does not take into account the connectivity of the immediate neighbours of the node under consideration. A number of centrality measures have been introduced that take into account the global connectivity properties of the network. These include various types of eigenvector centrality for both directed and undirected networks, Katz centrality, subgraph centrality and PageRank centrality [1]. The use of centrality scores provides rankings of the nodes in the networks. The higher the ranking of a node, the more important the node is believed to be within the network. There are many different ranking methods in use, and many algorithms have been developed to compute these rankings.

The purpose of this paper is to discuss some of the centrality measures and analyse the relationship between degree centrality, eigenvector centrality, and Katz centrality and to discuss the measures of centrality based on matrix functions including the logarithmic, sine, cosine, exponential and hyperbolic function. The main aim is to determine which of the matrix functions is highly correlated to “standard” centrality measures. We will use the Kendall correlation coefficient [2] in the experimental work to determine the correlations.

## 2. Literature Review

Bavelas [3] introduces the application of centrality of networks in human communication by measuring the communication within a small group in terms of the relationship between the structural centrality and influence in a group process. Afterwards, an application of centrality was made under the direction of Bavelas at the Group Networks Laboratory, M.I.T in the late 1940s. Leavitt in 1949 and Smith in 1950 conducted a study on centrality measure on which Bavelas in 1950 and Bavelas and Barrett in 1951 reported. These experiments all concluded that centrality was related to group efficiency in problem-solving and agreed with the subjective perception of leadership [4].

Various centrality measures in various contexts were then explored in the following decade. Cohn and Marriott in 1958 attempted to use the centrality to understand political integration in Indian social life [5]. Pitts examined the con-

sequences of centrality in communication paths for urban development [6]. Later, Czepiel used the concept of centrality to explain the pattern of diffusion of a technological innovation in the steel industry [7].

Recently, Bolland analysed the stability of the degree (DC), closeness (CC), betweenness (BC), and eigenvector (EC) centrality in random and systematic variations of network structure, and found that betweenness centrality changes significantly with the variation of the network structure, while degree and closeness centrality are usually stable. He also found that eigenvector centrality is the most stable of all the indices analysed [8]. Borgatti and Frantz extended the studies on the stability of centrality indices by considering the addition and deletion of nodes and links [9] as well as by differentiating several types of network topology such as uniformly random, small-world, core-periphery, scale-free, and cellular. Landherr reviewed critically the role of centrality measure in social networks [10]. Estrada analysed examples of how a particular centrality measure is applied in social networks [11].

Benzi and Klymko analysed centrality measures, such as degree, eigenvector, Katz and subgraph for both undirected and directed networks. They used the local and global influence of a given node in the network as measured by walks of different lengths through that node. They analysed the relationship between centrality measures based on the diagonal entries and row sums of the matrix exponential and resolvent of the adjacency matrix involving the degree and eigenvector centrality. They showed experimentally that the rankings produced by exponential subgraph centrality, total communicability and resolvent subgraph centrality converge to those produced by degree centrality [1].

Most of the centrality measures notations considered are combinatorial in nature and based on the discrete structure of the underlying networks. We can extend our studies by defining the centrality measure by using the spectral techniques from linear algebra. Benzi and Klymko considered the diagonal entries of the matrix exponential  $f(A) = e^A$  and the Katz function  $f(A) = (I - \alpha A)^{-1}$  where  $\alpha > 0$  and  $A$  is the adjacency matrix of the network [1]. Though, none of the previous studies considered other matrix functions such as the logarithmic, cosine, sine, hyperbolic functions and the generalized Katz centrality as centrality measures. In this work we will develop the notions of centrality based on matrix functions, and we will use the Kendall correlation coefficient [2] to determine the agreement between the node rankings produced by these matrix functions and those produced by the standard centrality measures.

### 3. Elements of Graph

Graphs are discrete structures which consist of vertices connected by edges. A graph can be written as  $G = (V, E)$ , where  $V(G)$  is a non-empty set of vertices (also called nodes) and  $E(G)$  is a set of edges. An edge of the graph consists of two vertices associated with it, these two vertices are called endpoints.

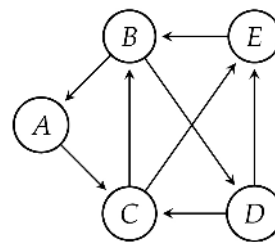
An edge starting and ending at the same vertex is called a loop.

- Graphs can be represented by using points as nodes (vertices) and joining them using line segments for edges. We write  $uv$  to denote an edge between nodes  $u$  and  $v$ .
- We can assign numerical values to the edges of a graph in which case the graph is referred to as *weighted*. In an *unweighted* graph we assign to every edge the value 1.
- If the edges of the graph are directed (**Figure 1**), then the graph is called a directed graph or digraph otherwise it is called an undirected graph. An undirected unweighted (**Figure 2**) graph without loops is *simple* if no two edges connect the same pair of vertices.
- For undirected graphs, if there are multiple edges between a pair of nodes then the graph is called a *multi-graph* or *pseudo-graph*. In digraphs we can have two edges connecting two vertices.

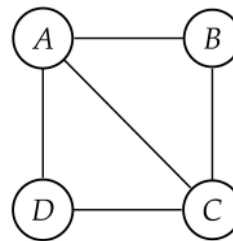
### 3.1. Basic Graph-Theoretic Terminology

If  $uv \in E$  is an edge in an undirected graph  $G$ , then nodes  $u$  and  $v$  are incident to the edge  $uv$  and we say that  $u$  and  $v$  are adjacent (neighbours) in  $G$ . It follows that  $u$  and  $v$  are the endpoints of an edge  $uv$ . If  $G$  is the directed graph and  $uv \in E$ , then  $u$  is said to be adjacent to  $v$  and  $v$  is said to be adjacent from  $u$ . We call  $u$  the initial node of  $uv$  and  $v$  the terminal or end node of  $uv$ . For an undirected graph  $G$ , the degree  $deg(v_i)$  of a node  $v_i$  is the total number of edges which are incident to the corresponding node  $v_i$ .

The degree  $deg(v_i)$  is the number of “immediate neighbours” of a node  $v_i$  in  $G$ . In a regular graph all nodes have the same degree. Nodes in directed graphs have an in-degree and an out-degree. The in-degree of a node  $v_i$ ,  $deg^-(v_i)$ , is the total number of edges with  $v_i$  as their terminal node. Similarly,



**Figure 1.** Digraph.



**Figure 2.** Undirected.

the out-degree of  $v_i$ ,  $\text{deg}^+(v_i)$ , is the total number of edges with  $v_i$  as their initial node. Loops contribute 1 to both the in-degree and out-degree.

A graph  $H = (W, F)$  is a subgraph of  $G = (V, E)$  if  $W \subseteq V$  and  $F \subseteq E$ . We write  $H \subseteq G$ , meaning that  $H$  is contained in  $G$  (or  $G$  contains  $H$ ). If  $H$  contains all edges of  $G$  that join two vertices in  $W$ , then we say that  $H$  is a subgraph induced or spanned by  $W$ .

The subgraphs of  $G = (V, E)$  can be obtained by deleting edges and vertices of  $G$ . We denote by  $G - W \equiv \langle V \setminus W \rangle$  with  $W \subset V$ , the subgraph of  $G$  obtained by deleting the vertices in  $W$  and all the edges incident with those vertices. In a similar manner, we denote by  $G - F \equiv (V, E \setminus F)$  where  $F \subset E$ , the subgraph of  $G$  obtained by deleting all the edges in  $F$ .

### 3.2. Walk, Trail and Path

A *walk*  $W_k$  of length  $k$  from node  $v_0$  to node  $v_k$  is a finite sequence  $W_k = (V, E)$  of the form

$$V = \{v_0, v_1, \dots, v_k\},$$

$$E = \{(v_0, v_1), (v_1, v_2), \dots, (v_{k-1}, v_k)\}$$

where  $v_i$  and  $v_{i+1}$  are adjacent.

A *trail* in  $G$  is a walk in which no edges of  $G$  appear more than once (a walk with all different edges). A trail which begins and ends at the same node is known as a closed trail or circuit.

A *path* in  $G$  is a walk in which no nodes appear more than once with the exception that  $v_k$  can be equal to  $v_0$ . A *cycle* or a *closed path* is a path which begins and ends at the same node.

Two nodes  $v_i$  and  $v_j$  in  $G$  are connected if there is a path between them. We say that graph  $G$  is connected if for every pair of nodes  $v_i$  and  $v_j$  there exists a path that starts at  $v_i$  and ends at  $v_j$ . An edge  $v_i v_j \in E$  in a connected graph  $G = (V, E)$  is a bridge if  $G$  becomes disconnected if  $v_i v_j$  is deleted. If  $v_i$  and  $v_j$  are nodes of the directed graph  $G$ , then  $G$  is said to be strongly connected if for any two nodes  $v_i$  and  $v_j$  we can find a path from  $v_i$  to  $v_j$  and from  $v_j$  to  $v_i$ . It is weakly connected if there is a path between every two nodes in the underlying undirected graph. The undirected graph is obtained by ignoring the directions of the edges in the directed graph. All strongly connected directed graphs are also weakly connected.

The digraph in **Figure 3** is strongly connected because there is a path between any two ordered vertices in the directed graph. The digraph in **Figure 4** is not strongly connected, since, for example, there is no directed path from  $S$  to  $T$ , nor from  $V$  to  $T$ , but it is weakly connected.

## 4. Matrices in Graphs

This section discusses the way of representing graphs using matrices. There are multiple ways to do this and any graph  $G(V, E)$  can be represented by using either adjacency, incidence or Laplacian matrices.

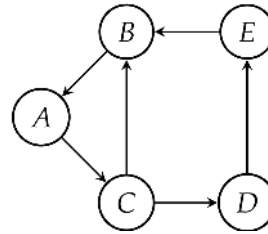


Figure 3. Strongly connected.

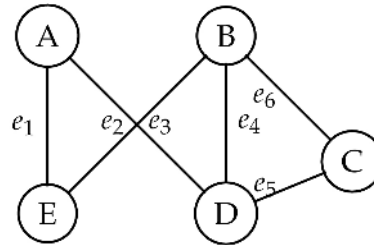


Figure 4. Weakly connected.

### 4.1. Matrices for Undirected Graph

Given an undirected graph  $G$  with  $n$  vertices and  $m$  edges. The *adjacency matrix* of  $G(V, E)$  is given by  $A = \{a_{ij}\}_{n \times n}$ , where

$$a_{ij} = \begin{cases} 1 & \text{if node } v_i \text{ is adjacent to node } v_j \\ 0 & \text{otherwise} \end{cases}$$

The *incidence matrix* is given by  $M = \{m_{i,j}\}_{n \times m}$ , where

$$m_{ij} = \begin{cases} 1 & \text{when edge } e_j \text{ is incident with node } v_i \\ 0 & \text{otherwise} \end{cases}$$

The *Laplacian matrix* can be found by using the relation

$$L = D - A,$$

where  $D$  is a diagonal matrix whose  $i$ th diagonal entry is the degree of the  $i$ th node and  $A$  is the adjacency matrix.

In other words,  $L = \{l_{ij}\}_{n \times n}$ , where

$$l_{ij} = \begin{cases} -1 & \text{if node } v_i \text{ is adjacent to node } v_j \\ \text{deg}(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Figure 5 is undirected graph with five nodes and six edges where there is a path from one node to the other nodes. The adjacency matrix  $A$ , the incidence matrix  $M$  and the diagonal matrix  $D$  will be

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

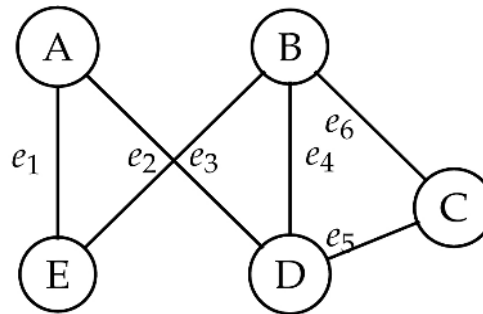


Figure 5. Undirected Graph.

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

The Laplacian matrix will be

$$L = D - A = \begin{pmatrix} 2 & 0 & 0 & -1 & -1 \\ 0 & 3 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ -1 & -1 & -1 & 3 & 0 \\ -1 & -1 & 0 & 0 & 2 \end{pmatrix}.$$

#### 4.2. Matrices for Directed Graph

Given a directed graph  $G$  with  $n$  nodes and  $m$  edges. The *adjacency matrix* of  $G(V, E)$  is given by  $A = \{a_{ij}\}_{n \times n}$ , where

$$a_{ij} = \begin{cases} 1 & \text{if node } v_i \text{ is connected to node } v_j \text{ and directed from } v_i \text{ to } v_j \\ -1 & \text{if node } v_i \text{ is connected to node } v_j \text{ and directed from } v_j \text{ to } v_i \\ 0 & \text{otherwise} \end{cases}$$

The *incidence matrix* is given by  $M = \{m_{i,j}\}_{n \times m}$ , where

$$m_{ij} = \begin{cases} 1 & \text{if node } v_i \text{ is the starting node of an edge } e_j \\ -1 & \text{if node } v_i \text{ is the end node of an edge } e_j \\ 0 & \text{otherwise} \end{cases}$$

Figure 6 is a digraph which shows that there is path from node  $A$  to any other node of the graph, nevertheless there is no path from node  $C$  to any other node of the network. The adjacency matrix  $A$  and the incidence matrix  $M$  will be

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ -1 & 0 & 1 & -1 & 1 \\ 0 & -1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{pmatrix},$$

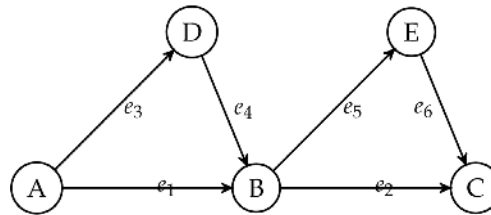


Figure 6. Directed Graph.

$$M = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

### 4.3. Distance in Graphs

Geodesic distance denoted as  $d(v_i, v_j)$  between two nodes  $v_i$  and  $v_j$  in the graph  $G$  is defined as the length of the shortest path between nodes  $v_i$  and  $v_j$ . The diameter of the graph  $G = (V, E)$  is given as

$$diam(G) = \max_{v_i, v_j \in V} d(v_i, v_j).$$

Geodesic distances in digraph  $Q$  in Figure 3:

$$d(D, C) = 4, \quad d(A, E) = 3 \text{ and } d(D, D) = 0.$$

The distance matrix of a graph, denoted as  $D$ , is the square matrix  $\{d(ij)\}_{n \times n}$ , where  $d(ij)$  is equal to the length of the shortest path from node  $v_i$  to node  $v_j$ .

Consider the undirected unweighted graph in Figure 7 as example;

The distance matrix for the graph in Figure 7, is given as;

$$D = \begin{pmatrix} 0 & 1 & 2 & 2 & 2 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 & 2 & 2 \\ 2 & 1 & 1 & 0 & 1 & 2 \\ 2 & 1 & 2 & 1 & 0 & 1 \\ 1 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

### 4.4. Perron—Frobenius Theorem

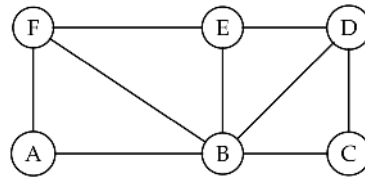
We will state (without giving the proof) the Perron-Frobenius theorem which will be used later on in our work.

**Theorem 1 (Perron-Frobenius theorem).** Let  $A \in \mathbb{R}^{n \times n}$  be an irreducible matrix. Then the Perron-Frobenius theorem states that [12]:

- $A$  has a principal eigenvalue  $\lambda_1$  such that all other eigenvalues  $\lambda_i$ , for  $i = 2, 3, \dots, n$ , satisfy

$$|\lambda_1| \geq |\lambda_i|.$$





**Figure 7.** Unweighted-Undirected Graph.

- The principal eigenvalue  $\lambda_1$  has algebraic and geometry multiplicity of 1, and has a right eigenvector  $\mathbf{x}$  with all positive elements, *i.e.*  $x_i > 0, \forall i = 1, 2, \dots, n$  and a left eigenvector  $\mathbf{v}$  with all positive elements, *i.e.*  $v_i = 0, \forall i = 1, 2, \dots, n$ .
- Any non-negative right eigenvector is a multiple of  $\mathbf{x}$ , any non-negative left eigenvector is a multiple of  $\mathbf{v}$ .

Furthermore, if  $A$  is the adjacency matrix of a directed network with a strongly connected component, then

- $A$  has a principal eigenvalue  $\lambda_1$  such that all other eigenvalues  $\lambda_i$ , for  $i = 2, 3, \dots, n$ , satisfy

$$|\lambda_1| \geq |\lambda_i|.$$

- The principal eigenvalue  $\lambda_1$  has algebraic and geometric multiplicity equal to 1, and has a left eigenvector  $\mathbf{x}$  with non-negative elements, *i.e.*  $x_i > 0$  if node  $i$  belongs to the strongly connected component of the network or the out-components of the network, and  $x_i = 0$  if node  $i$  belongs to the in-component of the strongly connected component of the network.

## 5. Centrality Measures

Centrality of a given node is a measure of the importance and influence of that node in the corresponding network. The identification of which nodes are more important or central than the others is a key issue in network analysis. We can ask the following questions:

- Which are the most central nodes in a network?
- Which are the most important nodes in a network?
- Which are the most influential nodes in a network?

These types of questions can have different interpretations in different networks. For instance;

- when dealing with a social network, the most central node can be the most popular person,
- when dealing with a web portal network, the most central node can be a web page with the best quality of content in a specific field,
- in terms of the internet network, the most central node might be a network gateway (router) with the highest bandwidth.

These ideas can be used to characterize types of centrality measure to find the most important nodes in a network in a given context. That is, there are many different centrality measures. When measuring the centrality of the node, we should be sure that:

- we know what each centrality measure means;

- what they measure well; and
- why a particular centrality measure is the most appropriate for the kind of set we are investigating.

The most common centrality measures include degree centrality, betweenness centrality, eigenvector centrality, Katz centrality, PageRank centrality, closeness centrality and subgraph centrality [11] [13].

### 5.1. Degree Centrality

*Degree centrality* of a node  $v_i$  in a given network is given by the total degree  $d_i$  of  $v_i$ . The degree centrality measures the ability of a node to communicate directly with other nodes.

In an undirected network, the degree  $d_i$  of a node is given as

$$d_i = \sum_{j=1}^n a_{ij} = \mathbf{e}_i^T \mathbf{A} \mathbf{e}, \tag{1}$$

where  $\mathbf{A}$  is the adjacency matrix,  $\mathbf{e}_i$  is the  $i$ th standard basis vector ( $i$ th column of the identity matrix) and  $\mathbf{e}$  is the vector of all entries one.

In a directed network, we can consider the in-degree of a node, given as

$$d_i^{in} = \sum_{j=1}^n a_{ji} = \mathbf{e}_i^T \mathbf{A} \mathbf{e}, \tag{2}$$

or the out-degree of the node, given as

$$d_i^{out} = \sum_{j=1}^n a_{ij} = \mathbf{e}^T \mathbf{A} \mathbf{e}_i. \tag{3}$$

In a directed graph, a source is a node with zero in-degree and a sink is a node with zero out-degree.

As an example, consider the undirected graph in **Figure 8**. We are interested in finding the central node using degree centrality.

The adjacency matrix  $\mathbf{A}$  for Network-1 in **Figure 8** is given as

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

The degree centralities are contained in the vector  $\mathbf{d} = \mathbf{A} \mathbf{e}$  ;

$$\mathbf{d} = [6 \ 3 \ 3 \ 4 \ 2 \ 3 \ 4 \ 2 \ 4 \ 3 \ 2 \ 2]^T$$

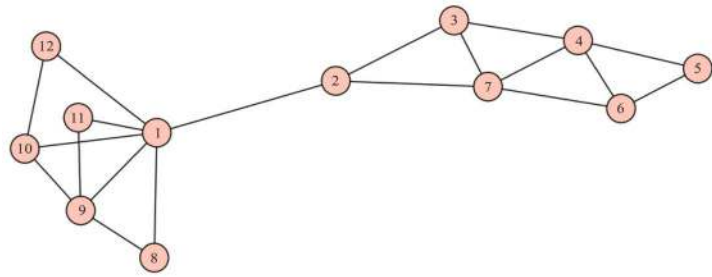


Figure 8. Network-1.

Using the degree centrality measure, node 1 is the most central node due to the fact that it has the highest degree.

## 5.2. Closeness Centrality

*Closeness centrality* measures the average shortest path length from a node to all other nodes. It uses neighbours and the neighbours of neighbours of a node  $v_i$  to determine its centrality. Thus, nodes that are not directly connected to  $v_i$  are taken into consideration as opposed to the degree centrality case.

Letting  $d(ij)$  be the length of the shortest path from node  $i$  to node  $j$ , the mean distance from node  $i$  to the other nodes in a network is given by:

$$l_i = \frac{1}{N-1} \sum_{i \neq j \in V} d(ij) = \frac{1}{N-1} \mathbf{e}_i^T \mathbf{D} \mathbf{e}, \quad (4)$$

where  $N$  is the total number of nodes and  $\mathbf{D}$  denotes the distance matrix.

In general, we want to associate high centrality score with important nodes. So we will use the reciprocal of  $l_i$  as the value of the centrality. Thus, the closeness centrality  $c_i$  for a node  $v_i$  is given by:

$$c_i = l_i^{-1} = \frac{N-1}{\mathbf{e}_i^T \mathbf{D} \mathbf{e}} \quad (5)$$

For example, consider Network-1 in **Figure 8**.

The distance matrix is given by  $\mathbf{D} = (d(i, j))$

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 2 & 1 & 2 & 2 & 2 & 2 & 2 \\ 2 & 1 & 0 & 1 & 2 & 2 & 1 & 3 & 3 & 3 & 3 & 3 \\ 3 & 2 & 1 & 0 & 1 & 1 & 1 & 4 & 4 & 4 & 4 & 4 \\ 4 & 3 & 2 & 1 & 0 & 1 & 2 & 5 & 5 & 5 & 5 & 5 \\ 3 & 2 & 2 & 1 & 1 & 0 & 1 & 4 & 4 & 4 & 4 & 4 \\ 2 & 1 & 1 & 1 & 2 & 1 & 0 & 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 0 & 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 1 & 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 & 2 & 1 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 2 & 0 & 2 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 2 & 1 & 2 & 0 \end{pmatrix}.$$

Then,

$$l_i = \frac{\mathbf{e}_i^T \mathbf{D} \mathbf{e}}{N-1} = \frac{\mathbf{e}_i^T \mathbf{d}}{N-1},$$

where  $\mathbf{D} \mathbf{e} = \mathbf{d}$

$$\mathbf{EVC} = [20 \ 20 \ 24 \ 29 \ 38 \ 30 \ 23 \ 29 \ 27 \ 28 \ 29 \ 29]$$

Thus,  $c_i = \frac{1}{l_i}$ , gives

$$\begin{aligned} c_1 &= \frac{11}{20} = 0.55, c_2 = \frac{11}{20} = 0.55, c_3 = \frac{11}{24} = 0.458 \\ c_4 &= \frac{11}{29} = 0.379, c_5 = \frac{11}{38} = 0.289, c_6 = \frac{11}{30} = 0.367 \\ c_7 &= \frac{11}{23} = 0.478, c_8 = \frac{11}{29} = 0.379, c_9 = \frac{11}{27} = 0.407 \\ c_{10} &= \frac{11}{28} = 0.393, c_{11} = \frac{11}{29} = 0.379, c_{12} = \frac{11}{29} = 0.379. \end{aligned}$$

Now, nodes 1 and 2 are identified as the most important nodes in the network.

### 5.3. Eigenvector Centrality

In an undirected network which is connected we can write the measure of centrality by using the eigenvector centrality. Eigenvector centrality takes into consideration the importance of neighbours of a node. In degree centrality a node is awarded one centrality point for each neighbour. Eigenvector centrality gives each node a score which is proportional to the sum of the score of its neighbours. In eigenvector centrality, a node is important if it is linked to other important nodes. The larger an entry on the node, the more important the node is considered to be.

From the Perron-Frobenius theorem, the eigenvector associated to the principal eigenvalue of the adjacency matrix  $\mathbf{A}$  is unique if the network is strongly connected. We define the centrality of a node iteratively by using the sum of its neighbours' centralities. We initially assume that a node  $j$  has centrality  $x_j^{(0)} = 1$ . Then we calculate a new iteration  $x_i^{(1)}$  as the sum of the centralities of  $i$ 's neighbours.

That is,

$$x_i^{(1)} = \sum_j a_{ij} x_j^{(0)}, \quad (6)$$

where  $a_{ij}$  are the entries of the adjacency matrix.

In matrix form we write this as:

$$\mathbf{x}^{(1)} = \mathbf{A} \mathbf{x}^{(0)}. \quad (7)$$

After  $k$ -steps, we have

$$\mathbf{x}^{(k)} = \mathbf{A}^{(k)} \mathbf{x}^{(0)}. \quad (8)$$

Note that the eigenvector centrality is defined as  $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)}$ .

We can write  $\mathbf{x}^{(0)}$  as a linear combination of eigenvectors  $v_i$  of the

adjacency matrix  $A$ , that is,

$$\mathbf{x}^{(0)} = \sum_i c_i v_i, \tag{9}$$

where the  $c_i$  are constants.

Then, from Equation (8), we have

$$\mathbf{x}^{(k)} = A^{(k)} \sum_i c_i v_i = \sum_i c_i A^{(k)} v_i = \sum_i c_i \lambda_i^k v_i = \sum_i c_i \lambda_i^k \left( \frac{\lambda_i}{\lambda_1} \right)^k v_i = \lambda_1^k \sum_i c_i \left( \frac{\lambda_i}{\lambda_1} \right)^k v_i. \tag{10}$$

Therefore,

$$\frac{\mathbf{x}^{(k)}}{\lambda_1^k} = \sum_i c_i \left( \frac{\lambda_i}{\lambda_1} \right)^k v_i, \tag{11}$$

where  $\lambda_i$  is the eigenvalue associated with the eigenvector  $v_i$  and  $\lambda_1$  is the principal eigenvalue.

Since  $\frac{\lambda_i}{\lambda_1} < 1$ , for  $i = 2, 3, \dots, n$ , we have;

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\mathbf{x}^{(k)}}{\lambda_1^k} &= \lim_{k \rightarrow \infty} \sum_i c_i \left( \frac{\lambda_i}{\lambda_1} \right)^k v_i = \sum_i c_i \lim_{k \rightarrow \infty} \left( \frac{\lambda_i}{\lambda_1} \right)^k v_i = c_1 v_1 \\ &\text{as } \lim_{k \rightarrow \infty} \left( \frac{\lambda_i}{\lambda_1} \right)^k \rightarrow 0, \forall i > 1. \end{aligned}$$

This implies that the limiting centralities are proportional to the principal eigenvector  $v_1$  of the adjacency matrix.

Therefore, in matrix form, the eigenvector centrality  $\mathbf{x}$  satisfies

$$A\mathbf{x} = \lambda_1 \mathbf{x} \Rightarrow \mathbf{x} = \frac{1}{\lambda_1} A\mathbf{x} \tag{12}$$

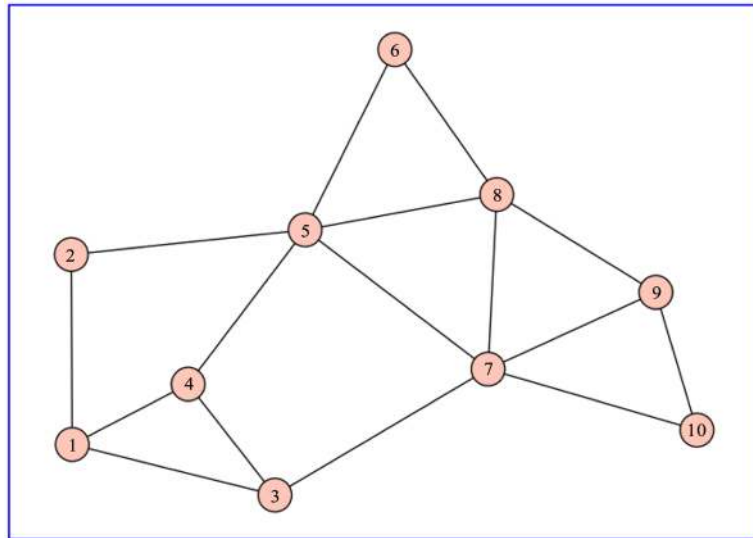
$$x_i = \frac{1}{\lambda_1} \sum_j a_{ij} x_j. \tag{13}$$

Note that in eigenvector centrality, the higher the centrality of the neighbours of the node, the more important the node is.

For example, consider the network in **Figure 9**.

The adjacency matrix for Network-2 in **Figure 9**, is given by

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$



**Figure 9.** Network-2.

The eigenvalues of the adjacency matrix are

$$\rho(A) = (-2.464, -1.931, -1.543, -0.666, -0.477, -0.332, 0.370, 1.398, 2.116, 3.531)$$

The principal eigenvector corresponding to the principal eigenvalue  $\lambda_1 = 3.531$  is

$$v = \begin{bmatrix} 0.198 & 0.181 & 0.261 & 0.255 & 0.443 \\ 0.243 & 0.468 & 0.416 & 0.313 & 0.221 \end{bmatrix}$$

Since the eigenvector centralities (*EVC*) correspond to the principal eigenvector, the eigenvector centralities for Network-2 will be

$$EVC = \begin{bmatrix} 0.198 & 0.181 & 0.261 & 0.255 & 0.443 \\ 0.243 & 0.468 & 0.416 & 0.313 & 0.221 \end{bmatrix}$$

Using the eigenvector centrality, we conclude that node 7 is the most important node.

### 5.4. Katz Centrality

Katz centrality takes into consideration both the number of direct neighbours and the further connections of a node in the network. That is, a node is important in Katz centrality if it has universal connections to other nodes in the network. Katz centrality takes into account all paths of arbitrary length from a node  $i$  to other nodes in the network.

The Katz centrality  $k$  is given by

$$k = \left( \sum_{k=0}^{\infty} \alpha^k A^k \right) e = \sum_{k=0}^{\infty} (\alpha A)^k e, \tag{14}$$

where  $e$  is the column vector of ones,  $\alpha$  is called the attenuation factor and  $A$  is the adjacency matrix of the network. We can expand Equation (14) as

$$\mathbf{k} = \sum_{k=0}^{\infty} (\alpha A)^k \mathbf{e} = \left( \mathbf{I} + \alpha A + (\alpha A)^2 + \dots \right) \mathbf{e}, \quad (15)$$

and, if the sum converges, then

$$\mathbf{k} = (\mathbf{I} - \alpha A)^{-1} \mathbf{e}, \quad (16)$$

where  $\mathbf{I}$  is  $n \times n$  identity matrix and  $\mathbf{e}$  is the column vector of ones.

To ensure the convergence of  $(\mathbf{I} - \alpha A)^{-1}$  and an accurate definition of Katz centrality, we must consider the attenuation factor  $\alpha$  to be within the range

$$\alpha \in \left( 0, \frac{1}{\lambda_1} \right).$$

For example, we compute the Katz centrality of Network-2 in **Figure 9**. Since the principal eigenvalue is  $\lambda_1 = 3.531$ , then  $\alpha \in \left( 0, \frac{1}{3.531} \right)$ . To calculate the Katz centrality, we choose  $\alpha = 0.25$ , then the Katz centrality will be

$$\begin{aligned} \mathbf{k} &= (\mathbf{I} - 0.25A)^{-1} \mathbf{e}, \\ \Rightarrow \mathbf{k} &= \begin{bmatrix} 5.826 & 5.223 & 7.086 & 6.992 & 11.065 \\ 6.316 & 11.526 & 10.201 & 7.895 & 5.855 \end{bmatrix}^T \end{aligned}$$

Therefore, node 7 is the most important node.

## 5.5. Subgraph Centrality

Subgraph centrality attempts to measure the centrality of a node by taking into consideration the participation of each node in all subgraphs of the network. It does this indirectly by counting the number of closed walks in the network which start and end at a given node in the network: a relationship can be shown between subgraphs and these walks.

If  $A$  is the adjacency matrix of an unweighted network, we know that

- $(A^k)_{ii}$  corresponds to the number of closed walks of length  $k$  starting at node  $v_i$ .
- $(A^k)_{ij}$  corresponds to the number of walks of length  $k$  that start at node  $v_i$  and end at node  $v_j$ .

We define

$$\mu_k(i) = (A^k)_{ii} \text{ as the local spectral moment of node } i$$

In a similar way to Katz centrality, *subgraph centrality* of a node  $i$  is a weighted sum of closed walks of different lengths which start and end at node  $i$ . The shorter the closed walk, the more the centrality of the node is influenced.

The subgraph centrality of node  $i$  in the network is given by

$$SC(i) = \sum_{k=0}^{\infty} \frac{\mu_k(i)}{k!} = \sum_{k=0}^{\infty} \frac{(A^k)_{ii}}{k!} = \left[ \mathbf{I} + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots \right]_{ii} \quad (17)$$

$$\Rightarrow SC(i) = (e^A)_{ii} \quad (18)$$

Considering the exponential of the adjacency matrix,

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots + \frac{A^k}{k!} + \dots$$

we observe that the numbers of closed walks associated with  $A^2$  of length 2 are counted twice for every link in the network. On the other hand, the closed walk associated with  $A^3$  of length 3 is counted  $2 \times 3 = 3!$  for every triangle. To avoid double counting, we penalize the closed walks of length 2 by  $2!$  and closed walks of length 3 are penalized by  $3!$ . In general, any circuit of length  $k$  can be traversed in 2 directions and there are  $k$  points where you can start, so that this circuit is counted  $2 \times k$  times. From the exponential of the adjacency matrix, when  $k \geq 4$ , the penalization included is not the same as the penalization of counting the number of repeated closed walks.

For instance, let us consider Network-2 in **Figure 9**. We want to find the subgraph centrality. We have;

$$e^A = \begin{pmatrix} 3.962 & 2.475 & 3.393 & 3.532 & 2.914 & 0.991 & 2.345 & 1.569 & 0.944 & 0.687 \\ 2.475 & 2.79 & 1.858 & 2.227 & 3.306 & 1.423 & 2.188 & 1.997 & 1.055 & 0.682 \\ 3.393 & 1.858 & 4.283 & 3.696 & 3.498 & 1.369 & 4.021 & 2.631 & 2.037 & 1.562 \\ 3.532 & 2.227 & 3.696 & 4.345 & 4.107 & 1.691 & 3.297 & 2.564 & 1.508 & 1.044 \\ 2.914 & 3.306 & 3.498 & 4.107 & 7.713 & 4.312 & 6.554 & 6.404 & 3.924 & 2.556 \\ 0.991 & 1.423 & 1.369 & 1.691 & 4.312 & 3.423 & 3.448 & 4.131 & 2.24 & 1.294 \\ 2.345 & 2.188 & 4.021 & 3.297 & 6.554 & 3.448 & 8.37 & 6.791 & 5.762 & 4.327 \\ 1.569 & 1.997 & 2.631 & 2.564 & 6.404 & 4.131 & 6.791 & 7.02 & 4.948 & 3.131 \\ 0.944 & 1.055 & 2.037 & 1.508 & 3.924 & 2.24 & 5.762 & 4.948 & 5.016 & 3.563 \\ 0.687 & 0.682 & 1.562 & 1.044 & 2.556 & 1.294 & 4.327 & 3.131 & 3.563 & 3.32 \end{pmatrix}.$$

Then the subgraph centrality will consist of the diagonal entries of  $e^A$ , which is

$$SC(i) = [3.962 \quad 2.79 \quad 4.283 \quad 4.345 \quad 7.713 \quad 3.423 \quad 8.37 \quad 7.02 \quad 5.016 \quad 3.32]^T$$

We observe that node 7 has the highest subgraph centrality, thus, node 7 is the most central node.

## 6. Relationship between Centrality Measures

Among the challenges that arise in determining the importance of a node in a network using centrality is that it is not always clear which of the centrality measures should be used. It is not obvious whether two centrality measures will give the same ranking of the nodes in the given network. Also, there is the necessity of choosing the attenuation factor  $\alpha$  in Katz centrality which adds another challenge. Different choices of  $\alpha$  may lead to different rankings. Experimentally, it has been seen that different centrality measures provide highly correlated rankings [1]. Ranking becomes more stable when  $\alpha$  approaches its limits, *i.e.* as



$$\alpha \rightarrow 0 \text{ and } \alpha \rightarrow \frac{1}{\lambda_1}.$$

We will prove these correlations and stability of ranking. This will relate the degree and eigenvector centrality to Katz centrality.

**Theorem 2.** *Let  $G=(V,E)$  be undirected connected network with adjacency matrix  $A$ . The Katz centrality  $k$  is given as*

$$k=(I-\alpha A)^{-1}e.$$

Then,

- as  $\alpha \rightarrow 0$ , the ranking produced by  $k$  converges to that produced by degree centrality, and
- as  $\alpha \rightarrow \frac{1}{\lambda_1}$ , the ranking produced by  $k$  converges to that produced by eigenvector centrality.

*Proof.* The Katz centrality  $k$  is given as

$$k=(I-\alpha A)^{-1}e, \tag{19}$$

which can be written as

$$\begin{aligned} k &= (I-\alpha A)^{-1}e \\ &= (I+\alpha A+(\alpha A)^2+(\alpha A)^3+\dots)e \end{aligned} \tag{20}$$

$$\begin{aligned} &= e+\alpha Ae+(\alpha A)^2e+(\alpha A)^3e+\dots \\ k &= e+\alpha d+(\alpha A)^2e+(\alpha A)^3e+\dots \end{aligned} \tag{21}$$

where  $d$  is the vector of the degree centralities of the nodes.

Consider the relation

$$\psi = \frac{1}{\alpha}(k-e). \tag{22}$$

It is clear that the ranking produced by  $\psi$  will be exactly the same as that produced by  $k$ , due to the fact that the score of each node has been scaled and shifted in the same way. Thus,

$$\begin{aligned} \psi &= \frac{1}{\alpha}(k-e) = \frac{1}{\alpha}(e+\alpha d+(\alpha A)^2e+(\alpha A)^3e+\dots) \\ \psi &= d+\alpha A^2e+\alpha^2A^3e+\dots. \end{aligned} \tag{23}$$

Then,

$$\lim_{\alpha \rightarrow 0} \psi = d \tag{24}$$

where  $d$  is the vector of the degree centralities of the nodes.

Therefore, the ranking produced by the Katz centrality reduces to that produced by degree centrality.

To show the second relation, we write the column vector  $e$ , as

$$e = \sum_{i=1}^n \beta_i v_i, \tag{25}$$

where  $\beta_i$ 's are constants and  $v_i$ 's are eigenvectors of matrix  $A$ .

Then, we can write the Katz centrality as

$$\begin{aligned} k &= (I - \alpha A)^{-1} e = \sum_{k=0}^{\infty} (\alpha A)^k e \\ &= \sum_{k=0}^{\infty} (\alpha A)^k \sum_{i=1}^n \beta_i v_i = \sum_{k=0}^{\infty} \alpha^k \sum_{i=1}^n \beta_i A^k v_i \\ &= \sum_{k=0}^{\infty} \alpha^k \sum_{i=1}^n \beta_i \lambda_i^k v_i = \sum_{i=1}^n \frac{1}{1 - \alpha \lambda_i} \beta_i v_i \end{aligned} \tag{26}$$

$$\begin{aligned} \Rightarrow k &= \frac{1}{1 - \alpha \lambda_1} \beta_1 v_1 + \sum_{i=2}^n \frac{1}{1 - \alpha \lambda_i} \beta_i v_i \\ &= \frac{1}{1 - \alpha \lambda_1} \left( \beta_1 v_1 + \sum_{i=2}^n \frac{1 - \alpha \lambda_1}{1 - \alpha \lambda_i} \beta_i v_i \right) \end{aligned} \tag{27}$$

Consider the relation

$$\phi = (1 - \alpha \lambda_1) k \tag{28}$$

That is, the ranking produced by  $\phi$  is exactly the same as that produced by  $k$ , due to the fact that the score of each node has been scaled and shifted in the same way. This implies that

$$\phi = \beta_1 v_1 + \sum_{i=2}^n \frac{1 - \alpha \lambda_1}{1 - \alpha \lambda_i} \beta_i v_i \tag{29}$$

Then,

$$\lim_{\alpha \rightarrow \frac{1}{\lambda_1}} \phi = \beta_1 v_1 \tag{30}$$

This implies that the limiting centralities are proportional to the principal eigenvector  $v_1$  of the adjacency matrix. Thus, the ranking produced by the Katz centrality reduces to those produced by eigenvector centrality.

### 7. Matrix Functions

This section discusses some of the matrix functions developed using Taylor series.

Matrix functions have applications throughout applied mathematics and scientific computing. Matrix functions are used in various fields, for example, in control theory and electromagnetism and can also be used to study complex networks like social networks.

Let  $f(z)$  be a complex-valued function, such that  $z \in C^{n \times n}$  and  $f(z)$  is analytic in the disc  $|z| < R$ , where  $R \in \mathbb{R}$ . Using Taylor's theorem, we can represent  $f(z)$  as a convergent power series

$$f(z) = a_0 + a_1 z + a_2 z^2 + \dots, \tag{31}$$

for  $|z| < R$ , and  $a_k, k \geq 0$  are complex-valued constants [12].

Let  $A \in C^{n \times n}$  be a complex-valued matrix. Then we define the matrix function of  $A$  as

$$f(A) = a_0 I + a_1 A + a_2 A^2 + \dots \quad (32)$$

The matrix series in Equation (32) is convergent to the  $n \times n$  matrix  $f(A)$  if all  $n^2$  scalar series that make up  $f(A)$  are convergent. It turns out that the series of  $f(A)$  converges if all eigenvalues of  $A$  lie in the region of convergence of  $f(z)$  in Equation (31). This can be proved by the following theorem.

**Theorem 3.** Suppose that  $f(z)$  has a power series representation, written as

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \quad (33)$$

in an open disc  $|z| < R$ . Then the series

$$\sum_{k=0}^{\infty} a_k A^k \quad (34)$$

is convergent if and only if the eigenvalues of  $A$  lie in  $|z| < R$  [12].

*Proof.* We prove this theorem only for diagonalisable matrices using the Jordan form of matrix  $A$ .

Let  $Q$  be a transformation matrix which diagonalizes  $A$ . Then we can write

$$D = \text{diag}(\lambda_1, \dots, \lambda_n) = Q^{-1} A Q. \quad (35)$$

Thus,

$$\begin{aligned} f(A) &= Q \text{diag}(f(\lambda_1), \dots, f(\lambda_n)) Q^{-1} \\ &= Q \text{diag}\left(\sum_{k=0}^{\infty} a_k \lambda_1^k, \dots, \sum_{k=0}^{\infty} a_k \lambda_n^k\right) Q^{-1} \\ &= Q \left(\sum_{k=0}^{\infty} a_k D^k\right) Q^{-1} \\ &= \sum_{k=0}^{\infty} a_k (Q D Q^{-1})^k \\ &= \sum_{k=0}^{\infty} a_k A^k. \end{aligned} \quad (36)$$

If  $A$  has an eigenvalue  $|\lambda_i| \geq R$ , then the series in Equation (33) diverges when evaluated at  $\lambda_i$ . It follows that the series in Equation (34) also diverges. That is, if there exist eigenvalues of matrix  $A$  which fall outside  $|z| < R$ , then the series in Equation (34) diverges.

Therefore  $f(A)$  converges if and only if  $|\lambda_i| < R$ , where  $\lambda_i, 1 \leq i \leq n$  are the eigenvalues of matrix  $A$ .

In general, if the function  $f(z)$  can be expressed by using Taylor series and it converges in the disc  $|z| < R$  which contains the eigenvalues of  $A$ , then  $f(A)$  can be computed by substituting the matrix  $A$  for variable  $z$  in the function  $f(z)$ . For instance,

$$f(z) = \frac{1+z}{1-z} \Rightarrow f(A) = (I - A)^{-1} (I + A)$$

The most important matrix functions which can be expressed by using the Taylor series are the following:

- Exponential function

$$f(z) = e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{z^k}{k!} \quad (37)$$

$$\Rightarrow f(A) = e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots$$

$$f(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!} \quad (38)$$

- Cosine function

$$f(z) = \cos(z) = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} + \dots = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}$$

$$\Rightarrow f(A) = \cos(A) = I - \frac{A^2}{2!} + \frac{A^4}{4!} + \dots$$

$$\cos(A) = \sum_{k=0}^{\infty} \frac{(-1)^k A^{2k}}{(2k)!} \quad (39)$$

- Sine function

$$f(z) = \sin(z) = z - \frac{z^3}{3!} + \frac{z^5}{5!} + \dots = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!}$$

$$\Rightarrow f(A) = \sin(A) = A - \frac{A^3}{3!} + \frac{A^5}{5!} + \dots = \sum_{k=0}^{\infty} \frac{(-1)^k A^{2k+1}}{(2k+1)!} \quad (40)$$

- Logarithmic function

$$f(z) = \log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots = \sum_{k=1}^{\infty} \frac{(-1)^{(k+1)} z^k}{k} \quad (41)$$

$$\Rightarrow f(A) = \log(I+A) = A - \frac{A^2}{2} + \frac{A^3}{3} - \frac{A^4}{4} + \dots = \sum_{k=1}^{\infty} \frac{(-1)^{(k+1)} A^k}{k} \quad (42)$$

- Hyperbolic function

#### 1) sinh function

$$f(z) = \sinh(z) = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots = \sum_{k=0}^{\infty} \frac{z^{2k+1}}{(2k+1)!}$$

$$\Rightarrow f(A) = \sinh(A) = A + \frac{A^3}{3!} + \frac{A^5}{5!} + \dots = \sum_{k=0}^{\infty} \frac{A^{2k+1}}{(2k+1)!}$$

#### 2) cosh function

$$f(z) = \cosh(z) = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k)!}$$

$$\Rightarrow f(A) = \cosh(A) = I + \frac{A^2}{2!} + \frac{A^4}{4!} + \dots = \sum_{k=0}^{\infty} \frac{A^{2k}}{(2k)!}$$

Each of these functions can (in theory) be used to define a centrality measure on a network with adjacency matrix  $A$ . For example, to obtain the centralities

of all the nodes we can compute  $f(\mathbf{A})\mathbf{e}$ , where  $\mathbf{e}$  is the vector of ones, or  $\text{diag}(f(\mathbf{A}))$ .

We may need to be careful with these raw centrality measures as  $f(\mathbf{A})$  may contain negative (or even complex) entries. For instance, to compute the logarithmic function  $f(\mathbf{A})$ , we need to take care of the complex entries, since it is not possible to make ranking out of complex entries. To avoid the complex entries, we compute  $\log(\gamma\mathbf{I} + \mathbf{A})$ , where  $\gamma$  is a real constant chosen so that  $(\gamma\mathbf{I} + \mathbf{A})$  has positive eigenvalues. The constant  $\gamma$  differs for different networks.

We can also define centrality measures by applying analytic continuations of  $f(z)$  outside its radius of convergence.

Recall that, if the attenuation factor  $\alpha < \frac{1}{\lambda_1}$ , where  $\lambda_1$  is the principal eigenvalue of  $\mathbf{A}$ , then

$$\sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k = (\mathbf{I} - \alpha\mathbf{A})^{-1}.$$

But

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

is also defined for  $|\alpha| \geq \frac{1}{\lambda_1}$  as long as  $\alpha \neq \frac{1}{\lambda_i}$  where  $\lambda_i$  is an eigenvalue of  $\mathbf{A}$ . Then we can generalize Katz centrality by the following definition:

$$\mathbf{k} = \text{diag}(\mathbf{I} - \alpha\mathbf{A})^{-1} \text{ or } \mathbf{k} = (\mathbf{I} - \alpha\mathbf{A})^{-1} \mathbf{e} \text{ with } \alpha > \frac{1}{\rho(\mathbf{A})}$$

where  $\mathbf{e}$  is the column vector of ones and  $\rho(\mathbf{A})$  is the spectrum of  $\mathbf{A}$ .

To determine which of the matrix functions can be used to assess centrality in the network, we will do some experimental work on a variety of networks in the following section. We will perform the experimental work by making comparisons between the rankings based on the common centrality measures discussed in section V and the rankings based on these matrix functions.

## 8. Experimental Work and Discussion

In this section we aim to analyse experimentally the agreements between the centrality measures discussed in section V, and whether the matrix functions discussed in section VII can be used to determine the important nodes in a network. The experimental work will compare matrix functions to the common centrality measures.

A variety of techniques can be used to compute centrality measures (those discussed in section V) and matrix functions. To compute the exponential of a matrix, logarithmic of a matrix and other matrix functions we will use *SciPy* matrix functions [14]. In our new measures involving matrix functions and generalisations to Katz centrality, we will calculate centralities by using the diagonal entries of these functions. We will use the Kendall correlation

coefficient in our experiments to compare the agreement between centrality measures.

### 8.1. Correlation (Kendall, Pearson, Spearman) Coefficient

Correlation is a bivariate analysis that measures the strengths of association between two variables. The value of the correlation coefficient varies between 1 and  $-1$ . The positive correlation signifies that the ranks of both variables are increasing, while the negative correlation signifies that as the rank of one variable increases, the rank of the other variable is decreasing. The correlation coefficient between the two variables is said to be a perfect association if it lies between  $\pm 1$  [15]. The closer the value of the correlation coefficient to 1 or to  $-1$ , the stronger the relationship between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. In statistics, we usually measure the strengths of association by: Pearson correlation, Kendall rank correlation and Spearman correlation.

**The Kendall coefficient of correlation** is the measure of the degree of correspondence between two set of ranks given to the same set of objects. The Kendall coefficient is interpreted as the difference between the probability of these objects being in the same order and the probability of these objects being in a different order [2].

Let  $X$  and  $Y$  be two observations, such that  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  is the set of joint random variables of  $X$  and  $Y$ , respectively. The values  $(x_i)$  and  $(y_i)$  are all unique.

- Any pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$  is said to be *concordant* if

$$\frac{x_i - x_j}{y_i - y_j} > 0,$$

which implies that either  $x_i > x_j$  and  $y_i > y_j$  or  $x_i < x_j$  and  $y_i < y_j$ .

- The pair is said to be *discordant* if

$$\frac{x_i - x_j}{y_i - y_j} < 0,$$

which implies that either  $x_i > x_j$  and  $y_i < y_j$  or  $x_i < x_j$  and  $y_i > y_j$ .

- The pair is neither *concordant* nor *discordant* if  $x_i = x_j$  or  $y_i = y_j$ .

The *Kendall ( $\tau$ ) correlation coefficient* is defined as

$$\tau = \frac{N^\circ \text{ concordant pairs} - N^\circ \text{ discordant pairs}}{\frac{1}{2}n(n-1)}$$

The Kendall rank coefficient can be interpreted as follows: the values of  $\tau$  greater than zero show an agreement, being close to one indicates a strong agreement. On the other hand, values less than zero show a disagreement and those close to negative one indicate a strong disagreement [2]. Indeed, if all pairs are *concordant*, then  $\tau = 1$ , which implies that the variables are in exactly the same ranking (order). If they are all *discordant* then  $\tau = -1$ , which implies that

the variables are in exactly the opposite ranking (order).

**Pearson correlation** is a measure of degree of linear relationship between two variables and is denoted by  $r$ . Pearson correlation is basically used to draw a line of best fit through the data of two variables,  $r$  indicates how far away all these data points are to the line of best fit.

The following formula is used to calculate the Pearson  $r$  correlation:

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{(N\sum x^2 - (\sum x)^2)(N\sum y^2 - (\sum y)^2)}} \quad (43)$$

where:

$r$  = Pearson correlation coefficient

$N$  = number of observation in each data set

$\sum xy$  = the sum of the products of paired scores

$\sum x$  = sum of scores of variable  $x$

$\sum y$  = sum of scores of variable  $y$  scores

$\sum x^2$  = sum of squared scores of variable  $x$

$\sum y^2$  = sum of squared scores of variable  $y$

To use Pearson correlation  $r$ , the two variables must be measured either in interval or ratio scale. However, both variables do not need to be measured on the same scale (for instance, one variable can be ratio and one can be interval). We can not use the Pearson correlation for ordinal data, instead we use Spearman's rank correlation or a Kendall's Correlation.

**Spearman rank correlation** is the nonparametric version of the Pearson correlation coefficient that is used to measure the degree of association between two two continuous or ordinal variables.

The following formula is used to calculate the Spearman rank correlation:

$$\rho = 1 - \frac{6\sum D_i^2}{N(N^2 - 1)} \quad (44)$$

where;  $D_i = R_{x_i} - R_{y_i}$  is the difference between the two ranks of each observation;  $N$  is the number of observations.

We use the Spearman correlation coefficient when the relationship between variables is not linear.

Despite the fact that both Spearman and Kendall correlations measure monotonicity relationships and have a nice interpretation but in this paper we will opt to use the Kendall correlation coefficient due to the following reasons [16] [17]:

- The distribution of Kendall's has better statistical properties.
- The interpretation of Kendall's in terms of the probabilities of observing the agreeable (concordant) and non-agreeable (discordant) pairs is very direct.
- The Kendall correlation has a smaller gross error sensitivity (GES) (more robust) and a smaller asymptotic variance (AV) (more efficient), that is Kendall correlation has a  $O(n^2)$  computation complexity comparing with

$O(n \log n)$  of Spearman correlation, where  $n$  is the sample size.

## 8.2. Network Models

Networks have been around us for so many years and the study is not new. Graph theorists and mathematicians have been surrounded by problems where they were trying to make sense of these complex networks. As a result of this, random network theory was generated stating that nodes and links in a graph are connected randomly to each other. In this paper we will consider three networks model due to its significance:

- **Erdős-Rényi model:** are formed by completely random interactions between the nodes. Each node chooses its neighbours at random, constrained either by an overall number of relationships that might be assigned in the graph, or a probability of connecting to a certain neighbour [18]. Mathematically, each network would be following a poisson distribution. This distribution is such that vast majority of nodes have equal number of links and it is almost impossible to find outliers.
- **Barabási-Albert model:** these are scale-free networks which are formed by two simple mechanisms, growth and preferential attachment. The main prediction which a scale free network makes is the presence of few outlier nodes which have many connections. These nodes are also known as hubs. Preferential attachment is a probabilistic mechanism in which a new node is free to connect to any node in the network, whether it is a hub or has a single link [18] [19].
- **Watts-Strogatz model:** is important because it shows how the “small-world effect” in networks can coexist with other commonly observed features of social networks, like a high clustering coefficient. More specifically, the model showed how adding a small fraction of random long-range links in an otherwise regular network can lead to slow, logarithmic scaling of the typical distance between nodes with network size [20].

### 8.2.1. First Experiment

We begin our experiments by considering a small network with 20 nodes. The network was randomly generated in text editor and drawn using Sage. The aim is to determine which of the matrix functions give similar rankings as the common centrality measures. We have many functions to choose from and we want to limit our choice. Note that we will not consider the exponential functions since it is similar to the subgraph centrality.

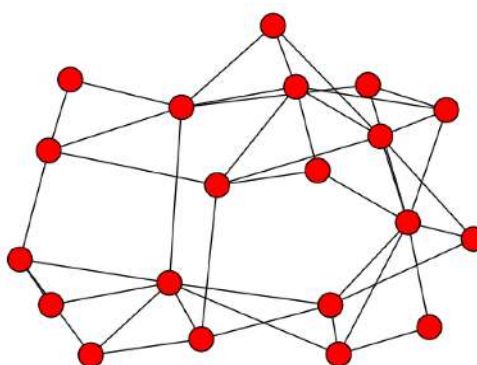
The experiment shows that the diagonal entries of the logarithmic function and cosine function give the ranking of the nodes in reverse order as compared to other rankings. Also, we observe that the sine function does not match any other centrality measure. The network in **Figure 10**, having 20 nodes and 42 edges gives us a real picture on node ranking.

The rankings of the nodes obtained for the graph in **Figure 10** by using different centrality measures including matrix functions are shown in **Table 1**.



**Table 1.** Rankings of nodes using centrality measures and matrix functions for the network in **Figure 10**.

Ranking	Centrality measure					Matrix functions					
	(l) 2-11	CC	DC	EC	KC	SC	log	cosh	sinh	cos	sin
1	6	19	19	19	19	19	8	19	19	16	9
2	12	12	3	3	3	3	13	3	3	13	10
3	19	3	4	4	12	16	12	4	8	8	8
4	2	4	12	12	4	11	4	12	5	16	16
5	14	6	14	14	14	14	7	6	14	18	13
6	3	2	0	6	6	6	18	14	0	7	11
7	4	14	6	2	0	1	2	17	11	6	6
8	15	0	2	0	2	5	0	2	1	5	5
9	17	9	17	17	17	17	17	17	6	0	1
10	1	10	18	18	15	15	0	15	18	17	12
11	5	15	5	15	18	10	10	15	15	17	17
12	9	17	15	5	10	15	9	7	10	2	2
13	7	1	7	7	7	9	5	10	14	0	0
14	18	5	1	1	5	14	18	11	9	18	18
15	0	7	9	9	9	2	7	5	3	3	3
16	10	11	10	10	1	4	1	1	4	7	7
17	11	18	16	11	11	3	11	9	2	19	19
18	16	8	11	16	16	6	16	16	6	15	15
19	8	13	8	13	13	12	13	13	12	14	14
20	13	16	13	8	8	19	8	8	19	4	4

**Figure 10.** Network with 20 nodes and 42 edges.

Note that the ranking is from the most to the least important/central node with respect to the centrality measure used.

To avoid making many comparisons using Kendall correlation coefficient between the common centrality measures and those produced by matrix

functions, we will choose one centrality measure among the other one. To do this we need to investigate whether the chosen centrality measure agrees with the other centralities measures. In this case, we make a comparison between the closeness centrality (CC) and degree centrality (DC), eigenvector centrality (EC), Katz centrality (KC) and subgraph centrality (SC) for graph in **Figure 10**.

**Table 2** shows that there is an agreement between closeness centrality and other centrality measures.

We observe from graph of **Figure 11** that there is an agreement between closeness centrality and other centrality measures.

In **Table 1**, we have to modify the cosine and logarithmic functions so that they match with other rankings. The best way of doing this seems to be by reversing the order of their rankings.

To be more confident about the rankings of nodes using matrix functions, we will use the Kendall correlation coefficient to make the comparison between closeness centrality and the matrix functions. We chose closeness centrality among the other standard centrality measure to make the comparison with matrix functions in as much as it takes into account neighbours and the neighbours of neighbours of a node to determine its centrality. In the comparisons, we denote by  $\tau(CC, f)$  the Kendall coefficient between closeness centrality and centrality measure induced by  $f(A)$ .

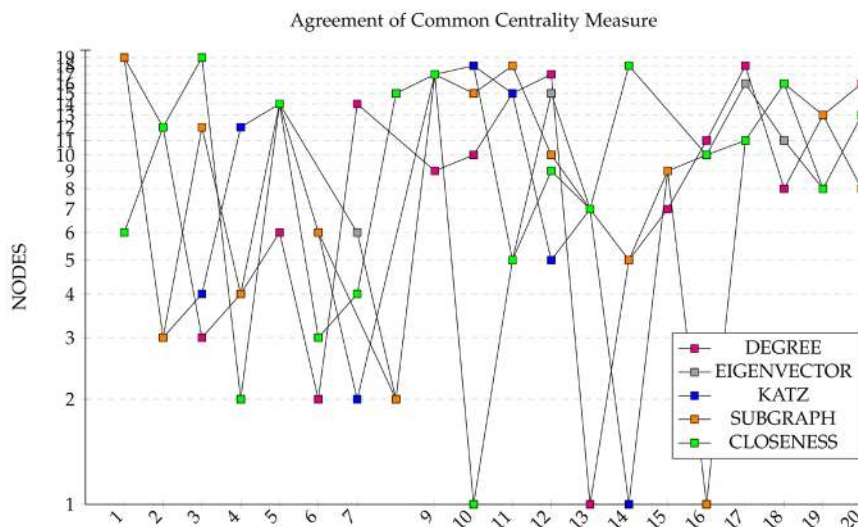
In **Table 3**, we reversed the rankings given by the cosine and the logarithmic functions before calculating the Kendall coefficients. We observe in **Table 3** that the Kendall coefficients  $\tau(CC, \log)$ ,  $\tau(CC, \cosh)$ ,  $\tau(CC, \sinh)$  and  $\tau(CC, \cos)$  are all positive. This implies the agreement of these matrix functions with the

**Table 2.** Kendall coefficients between closeness centrality and other centrality measure applied to graph in **Figure 10**.

Comparison	Kendal Coefficient
$\tau(CC, DC)$	0.695
$\tau(CC, EC)$	0.653
$\tau(CC, KC)$	0.684
$\tau(CC, SC)$	0.695

**Table 3.** Kendall coefficients between closeness centrality and matrix functions applied to graph in **Figure 10**.

Comparison	Kendal Coefficient
$\tau(CC, \log)$	0.737
$\tau(CC, \cosh)$	0.674
$\tau(CC, \sinh)$	0.558
$\tau(CC, \cos)$	0.695
$\tau(CC, \sin)$	-0.211



**Figure 11.** Agreement.

closeness centrality measure and the agreement is quite strong for logarithmic, cosh, and cosine functions since their Kendall coefficients are close to 1. On the other hand, the Kendall coefficient between closeness centrality and sine function is negative, which implies that there is no agreement between their rankings.

### 8.2.2. Second Experiment

We compare the agreement of centrality measures, by generating 10 random networks using the Barabási-Albert preferential attachment model.

The Barabási-Albert model is a simple scale-free random graph generator. The network begins with an initial set of  $m_0 \geq 2$  nodes. The degree of each node in the initial network should be at least 1, if not, the network will always end up being disconnected. New nodes are free to attach to an existing node in the network. At each step, a new node is created and connected to an existing node. Each new node is connected to  $m \leq m_0$  existing nodes with a probability that is proportional to the number of links that the existing nodes already have. To use this method, we specify the number of nodes in the network ( $n$ ) and the number of new nodes form as they appear ( $m$ ) in such a manner that nodes with higher degree have a higher chance of being selected for attachment [21].

The comparisons involve rankings of nodes using centrality measures such as closeness centrality, degree centrality, eigenvector centrality, Katz centrality and subgraph centrality. We use  $k = (I - \alpha A)^{-1} e$  in this experiment to compute Katz centrality. In all cases we take  $\alpha < \frac{1}{\lambda_1}$ . For each choice of  $n$  and  $m$ , we

generate 5 networks and record the mean values of the Kendall coefficients. We denote the Kendall coefficient correlation by  $\tau_i$  according to the **Table 4**.

It is evident from **Table 5**, that all Kendall coefficients are positive, which indicates an agreement between the rankings. The Kendall coefficient shows that

**Table 4.** The notations of Kendall correlation coefficients.

	DC	EC	KC	SC
CC	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$
DC		$\tau_5$	$\tau_6$	$\tau_7$
EC			$\tau_8$	$\tau_9$
KC				$\tau_{10}$

**Table 5.** Kendall coefficients for centrality measures applied to different random networks generated by using the Barabási-Albert method.

Networks	Kendall Coefficient									
	$(n, m)$	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$	$\tau_7$	$\tau_8$	$\tau_9$
(200, 2)	0.43	0.84	0.83	0.81	0.42	0.55	0.54	0.84	0.84	0.95
(200, 4)	0.59	0.88	0.89	0.88	0.57	0.60	0.57	0.96	0.99	0.96
(200, 5)	0.61	0.88	0.89	0.88	0.58	0.61	0.56	0.97	1.0	0.97
(400, 4)	0.55	0.89	0.89	0.89	0.54	0.56	0.54	0.98	1.0	0.98
(400, 7)	0.64	0.89	0.9	0.89	0.63	0.68	0.63	0.96	1.0	0.96
(600, 1)	0.19	0.82	0.59	0.79	0.18	0.43	0.24	0.64	0.85	0.74
(600, 16)	0.77	0.92	0.93	0.93	0.75	0.78	0.75	0.97	1.0	0.97
(700, 4)	0.45	0.90	0.91	0.91	0.44	0.48	0.44	0.97	0.99	0.97
(800, 10)	0.66	0.92	0.88	0.92	0.65	0.77	0.65	0.88	1.0	0.88
(800, 20)	0.78	0.93	0.91	0.93	0.77	0.85	0.77	0.92	1.0	0.92

$n$ : Number of nodes in the network.  $m$ : Number of neighbours to attach to each new node.

closeness centrality is highly correlated with centrality measures corresponding to  $\tau_2, \tau_3$  and  $\tau_4$ . The eigenvector, Katz and subgraph centralities are also highly related, as indicated by  $\tau_8, \tau_9$  and  $\tau_{10}$ . The experiment shows that the agreement becomes stronger if the network is more connected. In general, we say that for sufficiently dense (*i.e.*, very connected) networks, the two measures provide almost identical rankings, producing Kendall correlation coefficients close to 1.

### 8.2.3. Third Experiment

We generate 10 random networks as in the second experiment. This time, we fix the value of  $n$  to be 200 and we vary  $m$ . We calculate the Katz centrality for each network using different choices of  $\alpha$ . Recall, the Katz centrality of the nodes is given by  $\mathbf{k} = (\mathbf{I} - \alpha_i \mathbf{A})^{-1} \mathbf{e}$ . We choose

$$\alpha_1 = \frac{0.1}{\lambda_{\max}} \text{ for } \mathbf{k}_1, \alpha_2 = \frac{0.8}{\lambda_{\max}} \text{ for } \mathbf{k}_2,$$

$$\alpha_3 = \frac{1.5}{\lambda_{\max}} \text{ for } \mathbf{k}_3, \text{ and } \alpha_4 = \frac{10}{\lambda_{\max}} \text{ for } \mathbf{k}_4.$$

We calculate the Kendall correlation coefficients  $\tau(\mathbf{k}_i, \mathbf{k}_j)$  between all pairs

of measures as denoted in **Table 6**.

Note that the Kendall coefficient involving  $k_3$  and  $k_4$  are computed by taking the absolute value of the inverse function, that is  $\left| (I - \alpha A)^{-1} \right| e$  and the other coefficient are computed by using the normal formula.

We observe in **Table 7** that the Kendall coefficient between  $k_3$  and  $k_4$  is exactly 1. The Kendall coefficient between  $k_1$  and both  $k_3$  and  $k_4$ ,  $k_2$  and both  $k_3$  and  $k_4$  are the same. Since  $k_3$  and  $k_4$  correspond to the generalised Katz centrality (*i.e.*  $\alpha \geq \frac{1}{\lambda_{\max}}$ ), then we conjecture that the rankings provided by any generalised Katz centrality are always the same regardless of the choice of  $\alpha$ , this is when  $\alpha \geq \frac{1}{\lambda_1}$ .

### 8.2.4. Fourth Experiment

We repeat the second experiment involving generating 10 random networks with different number of nodes by using the Erdős-Rényi method. We will use the same notation for Kendall coefficients as we used in the second experiment, see **Table 4**.

The Erdős-Rényi model is used to generate random networks in which edges are set between nodes with equal probabilities. The model can be used to prove

**Table 6.** The notations of Kendall correlation coefficients.

	$K_2$	$K_3$	$K_4$
$K_1$	$\tau_{k_1 k_2}$	$\tau_{k_1 k_3}$	$\tau_{k_1 k_4}$
$K_2$		$\tau_{k_2 k_3}$	$\tau_{k_2 k_4}$
$K_3$			$\tau_{k_3 k_4}$

**Table 7.** Kendall coefficients for generalized Katz centrality applied to random networks generated by using the Barabási-Albert method.

Networks ( $n, m$ )	Kendall Correlation coefficient			
	$\tau_{k_1 k_2}$	$\tau_{k_1 k_3}$ & $\tau_{k_1 k_4}$	$\tau_{k_2 k_3}$ & $\tau_{k_2 k_4}$	$\tau_{k_3 k_4}$
(200, 1)	0.701	0.385	0.401	1.0
(200, 2)	0.703	0.526	0.493	1.0
(200, 3)	0.684	0.549	0.493	1.0
(200, 5)	0.734	0.638	0.639	1.0
(200, 6)	0.784	0.666	0.658	1.0
(200, 7)	0.777	0.697	0.677	1.0
(200, 8)	0.809	0.712	0.703	1.0
(200, 10)	0.828	0.759	0.761	1.0
(200, 15)	0.881	0.739	0.74	1.0
(200, 40)	0.911	0.479	0.54	1.0

the existence of networks satisfying various properties, it can also be used to provide a rigorous definition of what it means for a property to hold for almost all networks [22]. To generate random networks using the Erdős-Rényi model, we need to specify two parameters: the number of nodes in the network denoted by  $n$  and the probability  $p$  that a link should be formed between any two nodes [22].

The Kendall coefficients in **Table 8** are all positive and they are close to 1. Using the concept of Kendall coefficients, we say that rankings of nodes using centrality measures for random networks generated by using the Erdős-Rényi are highly correlated. This implies that there is a strong agreement in their rankings.

We repeat the third experiment but this time, generating 10 random networks with 200 nodes by using the Erdős-Rényi method. We will use the same definition of Katz centrality and the same notation as in **Table 6**.

**Table 9** shows that there is a high agreement between the rankings of  $k_1$  and  $k_2$ . On the other hand, we observe that the rankings of the nodes produced by Katz centrality when  $\alpha < \frac{1}{\lambda_1}$  and those produced by generalised Katz centrality, when  $\alpha \geq \frac{1}{\lambda_1}$  disagree, as we see from **Table 9** that the Kendall coefficient  $\tau_{k_1 k_3}$  and  $\tau_{k_2 k_3}$  are approximately zero.

### 8.2.5. Fifth Experiment

We repeat the second experiment, but now with 10 random networks generated by using the Watts-Strogatz method. The same notation for the Kendall coefficient will be used as that used in the second experiment, see **Table 4**.

**Table 8.** Kendall coefficients for centrality measures applied to different random networks generated by using the Erdős-Rényi method.

Networks ( $n, p$ )	Kendall Coefficient									
	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$	$\tau_7$	$\tau_8$	$\tau_9$	$\tau_{10}$
(200, 0.01)	0.668	0.836	0.847	0.779	0.593	0.785	0.843	0.767	0.691	0.921
(200, 0.04)	0.798	0.875	0.888	0.866	0.79	0.817	0.815	0.97	0.96	0.973
(200, 0.05)	0.804	0.864	0.876	0.861	0.783	0.806	0.79	0.973	0.986	0.979
(400, 0.04)	0.811	0.827	0.885	0.891	0.885	0.837	0.861	0.837	0.974	0.999
(400, 0.07)	0.833	0.841	0.849	0.841	0.882	0.905	0.882	0.976	1.0	0.976
(600, 0.02)	0.826	0.904	0.911	0.903	0.822	0.84	0.826	0.981	0.993	0.984
(600, 0.06)	0.842	0.853	0.859	0.853	0.892	0.919	0.892	0.972	1.0	0.972
(700, 0.05)	0.851	0.87	0.874	0.87	0.893	0.921	0.893	0.97	1.0	0.97
(800, 0.08)	0.884	0.866	0.877	0.866	0.919	0.944	0.919	0.974	1.0	0.974
(1000, 0.1)	0.997	0.937	0.942	0.937	0.937	0.942	0.937	0.995	1.0	0.995

$n$ : Number of nodes in the network.  $p$ : Probability of edge creation.

**Table 9.** Kendall coefficients for generalised Katz centrality applied to 10 random networks with 200 nodes generated by using the Erdős-Rényi method.

Networks ( $n, p$ )	Kendall Correlation coefficient			
	$\tau_{k_1 k_2}$	$\tau_{k_1 k_3}$ & $\tau_{k_1 k_4}$	$\tau_{k_2 k_3}$ & $\tau_{k_2 k_4}$	$\tau_{k_3 k_4}$
(200, 0.01)	0.888	0.033	0.055	1.0
(200, 0.02)	0.872	0.042	0.031	1.0
(200, 0.03)	0.913	0.007	0.003	1.0
(200, 0.05)	0.924	-0.003	-0.001	1.0
(200, 0.06)	0.924	0.054	0.062	1.0
(200, 0.08)	0.942	-0.034	-0.046	1.0
(200, 0.1)	0.94	0.006	0.001	1.0
(200, 0.15)	0.959	0.001	-0.003	1.0
(200, 0.3)	0.981	-0.004	-0.003	1.0
(200, 0.9)	1.0	0.043	0.043	1.0

$n$ : Number of nodes in the network.  $p$ : Probability of edge creation.

The Watts-Strogatz model was developed as a way to impose a high clustering coefficient onto classical random graphs. It produces networks with a small-world property. To generate these networks, we use `watts_strogatz_graph( $n, k, p$ )` in Sage. Here,  $n$  denotes the number of nodes in the network which are arranged in a ring and connected to  $k$  nearest neighbours in the ring. Each node is considered independently and, with probability  $p$ , a link is added between the node and one of the other nodes in the network, chosen uniformly at random in accordance with experiments detailed in [1].

In our experiment, we varied  $n, k$  and  $p$  and in each case, five networks were created. The averages of the Kendall coefficients over these 5 networks for different centrality measures are given in **Table 10**. These Kendall coefficients are computed for the complete set of rankings. The Kendall coefficients show that the agreement between the centrality measures are much weaker than for the networks produced by Barabási-Albert and Erdős-Rényi methods. The experiment also shows that as the network becomes denser the correlation between measures becomes stronger.

We then repeat the third experiment using the Watts-Strogatz method. The same definition of Katz centrality and the same notation were used as in **Table 6**.

In **Table 11** we see a high agreement between the rankings of  $k_1$  and  $k_2$ . The rankings by  $k_3$  and  $k_4$  are exactly the same. We also see the same pattern as in **Table 9**, so the same conclusion apply.

### 8.2.6. Sixth Experiment

The aim of this experiment is to use the three methods (Barabási-Albert, Erdős-Rényi and Watts-Strogatz) for generating random networks and to use

**Table 10.** Kendall coefficients for centrality measures applied to random networks generated by using the Watts-Strogatz method.

Networks $(n, k, p)$	Kendall Coefficient									
	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$	$\tau_7$	$\tau_8$	$\tau_9$	$\tau_{10}$
(200, 2, 0.2)	0.118	0.583	0.331	0.301	0.057	0.434	0.446	0.122	0.093	0.943
(400, 3, 0.1)	0.231	0.252	0.169	0.176	0.198	0.332	0.377	0.24	0.186	0.871
(500, 4, 0.3)	0.367	0.519	0.504	0.226	0.43	0.62	0.63	0.723	0.496	0.686
(500, 6, 0.3)	0.433	0.543	0.547	0.191	0.535	0.726	0.583	0.753	0.52	0.614
(600, 7, 0.3)	0.406	0.605	0.557	0.208	0.553	0.736	0.623	0.77	0.521	0.626
(400, 3, 0.4)	0.143	0.275	0.202	0.17	0.065	0.594	0.65	0.123	0.114	0.902
(300, 6, 0.1)	0.149	0.201	0.281	-0.115	0.37	0.537	0.421	0.604	0.453	0.528
(700, 2, 0.2)	0.048	0.477	0.189	0.167	0.044	0.437	0.453	0.155	0.133	0.953
(800, 100, 0.4)	0.992	0.919	0.935	0.919	0.921	0.938	0.921	0.98	1.0	0.98
(800, 110, 0.5)	1.0	0.936	0.942	0.936	0.936	0.942	0.936	0.993	1.0	0.993

$n$ : Number of nodes in the network.  $k$ : Each node is connected to  $k$  nearest neighbours.  $p$ : Probability of rewiring each edge.

**Table 11.** Kendall coefficients for generalised Katz centrality applied to 10 random networks with 200 nodes generated by using the Watts-Strogatz method.

Networks $(n, k, p)$	Kendall Correlation coefficient			
	$\tau_{k_1 k_2}$	$\tau_{k_1 k_3}$ & $\tau_{k_1 k_4}$	$\tau_{k_2 k_3}$ & $\tau_{k_2 k_4}$	$\tau_{k_3 k_4}$
(200, 2, 0.1)	0.968	-0.08	-0.083	1.0
(200, 4, 0.1)	0.915	0.021	0.026	1.0
(200, 4, 0.3)	0.863	0.006	0.047	1.0
(200, 3, 0.5)	0.86	-0.02	-0.021	1.0
(200, 6, 0.3)	0.897	-0.043	-0.032	1.0
(200, 8, 0.4)	0.914	0.012	0.02	1.0
(200, 8, 0.5)	0.926	-0.022	-0.007	1.0
(200, 10, 0.2)	0.903	0.088	0.128	1.0
(200, 40, 0.3)	0.954	0.006	0.025	1.0
(200, 100, 0.1)	0.988	0.037	0.047	1.0

the Kendall correlation coefficient to see whether there will be an agreement between the closeness centrality and matrix functions, such as the logarithmic, cosh, sinh, cosine and sine functions. Using each method, we generate 10 random networks and for each network we create five networks in which the Kendall coefficient will be obtained by taking the average over these 5 created networks. The aim is to see whether the pattern observed in our first experiment is repeated.

**Table 12** shows that there is an agreement between closeness centrality and matrix functions (logarithmic, cosh and sinh) in their ranking measures for



**Table 12.** Kendall coefficients between closeness centrality and matrix functions for 10 random networks generated by using the Barabási-Albert method.

Networks	Kendall Correlation coefficient				
	$\tau(\text{CC}, \log)$	$\tau(\text{CC}, \cosh)$	$\tau(\text{CC}, \sinh)$	$\tau(\text{CC}, \cos)$	$\tau(\text{CC}, \sin)$
(200, 2)	0.527	0.786	0.793	0.071	0.202
(200, 6)	0.665	0.877	0.877	-0.05	-0.118
(300, 4)	0.55	0.867	0.866	0.05	-0.05
(300, 8)	0.721	0.916	0.916	0.015	-0.027
(500, 2)	0.538	0.82	0.827	0.019	0.031
(500, 10)	0.685	0.913	0.913	0.033	0.054
(700, 1)	0.364	0.535	0.216	-0.131	0.216
(800, 15)	0.734	0.923	0.923	0.002	-0.037
(800, 30)	0.852	0.912	0.912	0.041	0.034
(1000, 100)	0.969	0.92	0.92	0.121	0.054

networks generated by using the Barabási-Albert method. On the other hand, the agreement between closeness centrality and other matrix functions such as cosine and sine is weak.

We now generate 10 random networks by using the Erdős-Rényi and Watts-Strogatz methods.

**Table 13** shows that networks generated by using the Erdős-Rényi method agree in the ranking measures given by closeness centrality and matrix functions such as logarithmic, cosh and sinh.

**Table 14** shows that the agreement between closeness centrality and the matrix functions (cosh, sinh, cosine and sine) are weak in many cases. When we use the Watts-Strogatz method to generate the networks, the logarithmic function, gives a ranking similar to closeness centrality when  $m \geq 10$ .

In general, the agreement between closeness centrality and hyperbolic functions (cosh and sinh) is not strong for the networks generated by using the Watt-Strogatz methods. The agreement between closeness centrality and the cosine function as well as the sine function in networks generated by using Barabási-Albert, Erdős-Rényi and Watts-Strogatz methods are all weak. Among the tested matrix functions, the logarithmic function gives the best agreement of ranking with other centrality measures.

### 8.2.7. Real-World Network Experiments

In this experiment, we will study the Kendall correlation coefficient for real-world networks. The networks in this experiment come from a variety of sources, some of these data we have obtained from Gephi sample datasets [23] and others are found in Pajek Data Sets [24]. We will compare only some of the centrality measures (closeness, subgraph and Katz) and the logarithmic matrix function. Note that we are not interested in the meaning of each node within the

**Table 13.** Kendall coefficients between closeness centrality and matrix functions for 10 random network generated by using the Erdős-Rényi method.

Networks ( $n, p$ )	Kendall Correlation coefficients				
	$\tau(\text{CC, log})$	$\tau(\text{CC, cosh})$	$\tau(\text{CC, sinh})$	$\tau(\text{CC, cos})$	$\tau(\text{CC, sin})$
(200, 0.1)	0.824	0.826	0.826	-0.16	-0.105
(200, 0.3)	0.651	0.928	0.928	0.032	0.056
(300, 0.4)	0.385	0.94	0.94	0.06	-0.038
(400, 0.2)	0.821	0.929	0.929	0.008	-0.011
(500, 0.5)	0.176	0.968	0.968	0.003	0.041
(600, 0.09)	0.866	0.858	0.858	-0.13	-0.03
(600, 0.4)	0.528	0.967	0.967	0.009	-0.02
(800, 0.3)	0.607	0.965	0.965	-0.013	-0.038
(900, 0.08)	0.895	0.894	0.894	0.027	0.011
(1000, 0.2)	0.77	0.958	0.958	0.036	0.002

**Table 14.** Kendall coefficients between closeness centrality and matrix functions for 10 random networks generated by using the Watts-Strogatz method.

Networks ( $n, m, p$ )	Kendall Correlation coefficient				
	$\tau(\text{CC, log})$	$\tau(\text{CC, cosh})$	$\tau(\text{CC, sinh})$	$\tau(\text{CC, cos})$	$\tau(\text{CC, sin})$
(200, 3, 0.5)	0.26	0.26	-0.167	0.094	0.011
(200, 3, 0.8)	0.279	0.273	-0.048	0.109	-0.026
(300, 2, 0.4)	0.285	0.285	0.101	0.103	0.067
(400, 8, 0.2)	0.511	0.101	0.003	0.426	0.316
(500, 10, 0.5)	0.721	0.612	0.514	0.177	0.239
(600, 3, 0.09)	0.21	0.213	0.104	0.017	0.104
(600, 12, 0.4)	0.684	0.475	0.441	0.179	0.231
(800, 10, 0.4)	0.682	0.413	0.296	0.249	0.247
(900, 12, 0.2)	0.615	-0.003	-0.017	0.382	0.357
(1000, 20, 0.8)	0.842	0.889	0.889	-0.338	0.001

network, what we really want to know is whether there is an agreement between these centrality measures in real-world networks. We chose the logarithmic function over other matrix functions since it shows a high agreement with other centralities when applied to random networks.

We can clearly see from **Table 15** that all Kendall coefficient are positive. This implies that there is an agreement between the rankings of nodes. We also observe that the agreement between centrality measures (closeness, Katz, subgraph) and the logarithmic function is high, irrespective of the connectivity of the underlying network. In general, we can say that the logarithmic function is the best of the tried matrix functions and can be used as a centrality measure, since it gives a ranking similar to rankings of other centrality measures.

**Table 15.** Kendall coefficients for the logarithmic function and some standard centrality measures as applied to 10 real-world networks.

Networks	Nodes	Kendall Correlation coefficient					
		$\tau(\text{CC,SC})$	$\tau(\text{CC,KC})$	$\tau(\text{CC,log})$	$\tau(\text{SC,KC})$	$\tau(\text{SC,log})$	$\tau(\text{KC,log})$
<b>Flying Team</b>	51	0.233	0.095	0.173	0.153	0.555	0.062
<b>Dolphins</b>	62	0.715	0.797	0.696	0.87	0.824	0.768
<b>Lesmis</b>	77	0.641	0.612	0.591	0.9	0.683	0.755
<b>Sun-Juan-Sur</b>	79	0.168	0.259	0.128	0.39	0.713	0.326
<b>Polbooks</b>	105	0.312	0.359	0.451	0.924	0.635	0.696
<b>Word adjacencies</b>	112	0.889	0.889	0.742	0.978	0.779	0.8
<b>College Football</b>	115	0.156	0.23	0.335	0.898	0.256	0.338
<b>Cele2</b>	307	0.681	0.688	0.634	0.954	0.755	0.795
<b>Key Words</b>	831	0.759	0.769	0.667	0.968	0.759	0.783
<b>Netscience</b>	1589	0.298	0.581	0.069	0.693	0.727	0.447

## 9. Conclusions

In this work we examined the centrality measures such as closeness, degree, eigenvector, Katz and subgraph. We showed the relationship between the Katz centrality and eigenvector as well as degree centrality. We developed our notion of centrality measure by considering the rankings of nodes based on matrix functions such as logarithmic, cosine, sine, hyperbolic functions and the generalised Katz centrality. We showed experimentally by using various classes of graphs that the rankings of the nodes given by closeness, degree, eigenvector, Katz and subgraph centrality are highly correlated. Moreover, we showed experimentally that the rankings of nodes given by different choices of attenuation factor  $\alpha$  for the generalised Katz centrality, in which  $\alpha \geq \frac{1}{\lambda_1}$  where  $\lambda_1$  is the principal eigenvalue of the network adjacency matrix  $A$ , are exactly the same. In terms of matrix functions, the experiment shows that there is a high agreement between the rankings of nodes given by the logarithmic function and other common centrality measures discussed in Section V.

Similar results were found to hold for real-world networks: the rankings given by the logarithmic function and those given by closeness, Katz and subgraph centrality are highly correlated irrespective of the connectivity of the network. In general, we concluded that the logarithmic function, out of the matrix functions we have considered, is the best and can be used as a centrality measure.

In this work, we considered only the diagonal entries of the matrix functions with some modifications of the calculations of their Kendall correlation coefficients. We also found that the logarithmic function gives a relatively good ranking as compared to the rankings given by other centrality measures. We suggest that in the future, we can consider the row sums of these matrix functions with or without any modification and examine whether they give

similar rankings as other centrality measures. The paper did not analyse the significance and the uses of the centrality measures based on matrix and it has been left for future work.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Benzi, M. and Klymko, C. (2013) Total Communicability as a Centrality Measure. *Journal of Complex Networks*, **1**, 124-149. <https://doi.org/10.1093/comnet/cnt007>
- [2] Kendall, M.G. (1938) A New Measure of Rank Correlation. *Biometrika*, **30**, 81-93. <https://doi.org/10.1093/biomet/30.1-2.81>
- [3] Bavelas, A. (1948) A Mathematical Model for Group Structures. *Human Organization*, **7**, 16-30. <https://doi.org/10.17730/humo.7.3.f4033344851gl053>
- [4] Bavelas, A. and Barrett, D. (1951) An Experimental Approach to Organizational Communication. American Management Association, New York, 57-62.
- [5] Cohn, B.S. and Marriott, M. (1958) Networks and Centres of Integration in Indian Civilization. *Journal of Social Research*, **1**, 1-9.
- [6] Pitts, F.R. (1965) A Graph Theoretic Approach to Historical Geography. *The Professional Geographer*, **17**, 15-20. [https://doi.org/10.1111/j.0033-0124.1965.015\\_m.x](https://doi.org/10.1111/j.0033-0124.1965.015_m.x)
- [7] Czepiel, J.A. (1974) Word-of-Mouth Processes in the Diffusion of a Major Technological Innovation. *Journal of Marketing Research*, **11**, 172-180. <https://doi.org/10.2307/3150555>
- [8] Bolland, J.M. (1988) Sorting out Centrality: An Analysis of the Performance of Four Centrality Models in Real and Simulated Networks. *Social Networks*, **10**, 233-253. [https://doi.org/10.1016/0378-8733\(88\)90014-7](https://doi.org/10.1016/0378-8733(88)90014-7)
- [9] Borgatti, S.P. and Everett, M.G. (2006) A Graph-Theoretic Perspective on Centrality. *Social Networks*, **28**, 466-484. <https://doi.org/10.1016/j.socnet.2005.11.005>
- [10] Landherr, A., Friedl, B. and Heidemann, J. (2010) A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering*, **2**, 371-385. <https://doi.org/10.1007/s12599-010-0127-3>
- [11] Estrada, E. (2012) *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Oxford.
- [12] Higham, N.J. (2008) *Bibliography of Functions of Matrices: Theory and Computation*. SIAM. <https://doi.org/10.1137/1.9780898717778>
- [13] Boldi, P. and Vigna, S. (2014) Axioms for Centrality. *Internet Mathematics*, **10**, 222-262. <https://doi.org/10.1080/15427951.2013.865686>
- [14] Higham, N.J. and Deadman, E. (2016) *A Catalogue of Software for Matrix Functions*. Version 2.0.
- [15] Bethea, R.M. (2018) *Statistical Methods for Engineers and Scientists*. Routledge, Abingdon-on-Thames.
- [16] Fredricks, G.A. and Nelsen, R.B. (2007) On the Relationship between Spearman's Rho and Kendall's Tau for Pairs of Continuous Random Variables. *Journal of Statistical Planning and Inference*, **137**, 2143-2150. <https://doi.org/10.1016/j.jspi.2006.06.045>

- 
- [17] Xu, W., Hou, Y., Hung, Y.S. and Zou, Y. (2010) Comparison of Spearman's Rho and Kendall's Tau in Normal and Contaminated Normal Models.
- [18] Prettejohn, B.J., Berryman, M.J. and McDonnell, M.D. (2011) Methods for Generating Complex Networks with Selected Structural Properties for Simulations: A Review and Tutorial for Neuroscientists. *Frontiers in Computational Neuroscience*, **5**, 11. <https://doi.org/10.3389/fncom.2011.00011>
- [19] Bayati, M., Kim, J.H. and Saberi, A. (2010) A Sequential Algorithm for Generating Random Graphs. *Algorithmica*, **58**, 860-910. <https://doi.org/10.1007/s00453-009-9340-1>
- [20] Meij, N. (2010) Networks: An Introduction.
- [21] Albert, R. and Barabási, A.L. (2002) Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, **74**, 47-97. <https://doi.org/10.1103/RevModPhys.74.47>
- [22] Newman, M.E., Strogatz, S.H. and Watts, D.J. (2001) Random Graphs with Arbitrary Degree Distributions and Their Applications. *Physical Review E*, **64**, Article ID: 026118. <https://doi.org/10.1103/PhysRevE.64.026118>
- [23] Datasets, Wikipedia, the Free Encyclopedia. <https://wiki.gephi.org/index.php/Datasets>
- [24] Old Pajek Data Sets. <http://pajek.imfm.si/doku.php?id=data:pajek:vlado>