CENTRALITY METRICS IN GRAPH NODE NETWORKS

Team 53

Abstract

This is End Report of Our Project and the team members are:

- Himanshu Yadav
- Archisha Panda
- Nida Shahnoor

1 Introduction

In various network systems such as social networks, transportation networks, information networks, and biological networks, the identification of important nodes or entities plays a crucial role in understanding their dynamics and optimizing their functionality. Centrality metrics provide a quantitative measure to identify and evaluate the significance or influence of individual nodes within a network. By analyzing centrality metrics, researchers and practitioners can gain insights into the structural properties of networks and identify critical elements that have a significant impact on network dynamics.

The centrality metrics problem revolves around finding effective measures to quantify the importance of nodes within a network. The problem is particularly relevant when analyzing complex systems, where understanding the role and impact of individual nodes can lead to improved decision-making, resource allocation, and network optimization.

There are several popular centrality metrics that have been developed and widely studied to address this problem. These include degree centrality, betweenness centrality, eigenvector centrality, closeness centrality, and PageRank, among others. Each centrality metric offers a unique perspective on node importance, capturing different aspects of network dynamics and significance.

Solving the centrality metrics problem involves both theoretical analysis and practical computation. Linear algebraic techniques, graph algorithms, and network analysis methods are commonly employed to compute centrality metrics efficiently and accurately.

By addressing the centrality metrics problem, researchers aim to advance our understanding of network structures and their behaviour, enabling more effective analysis, optimization, and decision-making in diverse fields such as social sciences, transportation, communication networks, epidemiology, and infrastructure planning.

2 Computations of different centralities

In network analysis, centralities are used to determine a node's significance or centrality inside a network. There are various centrality metrics, each of which addresses a different aspect of centrality. Here are a few typical centrality measurements and how they are calculated:

2.1 Degree centrality

Degree Centrality:
Degree centrality is a concept in network analysis that measures the importance or prominence of a node within a network based on its degree, which is the number of connections it has to other nodes. It is one of the most basic and widely used centrality measures and provides a simple way to identify the most connected nodes in a network.
In a network, nodes represent individual entities (such as individuals, organizations, or websites), and edges represent the connections or relationships between them. Degree centrality focuses on the local structure of the network, specifically the number of direct connections a node has, regardless of the properties of those connections.
There are two types of degree centrality: indegree centrality and outdegree centrality.

1. **Indegree Centrality:** Indegree centrality measures the number of incoming connections to a node. It quantifies how many other nodes are directly linked to a particular node. In social networks, indegree centrality can represent measures such as popularity or influence. For example, in a Twitter network, a user with a high indegree centrality may receive a large number of mentions or retweets.

2. **Outdegree Centrality:** Outdegree centrality measures the number of outgoing connections from a node. It indicates from how many other nodes a particular node is directly connected to. Outdegree centrality can represent measures such as activity or expansiveness.
   For Undirected Network Indegree and Outdegree is same.
   It's Computation using linear algebra.
   Degree centrality can be computed using linear algebra by representing the network as an adjacency matrix. An adjacency matrix is a square matrix where the rows and columns represent nodes in the network, and the elements of the matrix indicate the presence or absence of edges between

nodes.

For Undirected Network

1. **Construct the Adjacency Matrix:**Create an N x N adjacency matrix A, where A[i][j] = 1 if there is an edge between nodes i and j, and A[i][j] = 0 otherwise. In an undirected network, the adjacency matrix is symmetric, i.e.,

$$\text{A[i][j]} = \text{A[j][i]}.$$

2. **Compute the Degree Vector:** Compute the degree vector d of length N, where d[i] represents the degree of node i. The degree of a node i is simply the sum of the elements in the i-th row (or column) of the adjacency matrix. Mathematically, $d[i] = \Sigma A[i][j]$ for j = 1 to N.

3. **Normalize the Degree Vector:** Optionally, you can normalize the degree vector to obtain a normalized degree centrality measure. The normalized degree centrality divides each element of the degree vector by N-1, the maximum possible degree in the network. This normalization scales the centrality values to the range [0, 1]. The normalized degree centrality for node i is given by

$$c[i] = \frac{d[i]}{(N-1)}$$

.

For example

Let's take this case.

Node 1 is connected to nodes 2 and 3.

Node 2 is connected to nodes 1, 3, and 4.

Node 3 is connected to nodes 1, 2, and 4.

Node 4 is connected to nodes 2, 3, and 5.

Node 5 is connected to node 4.

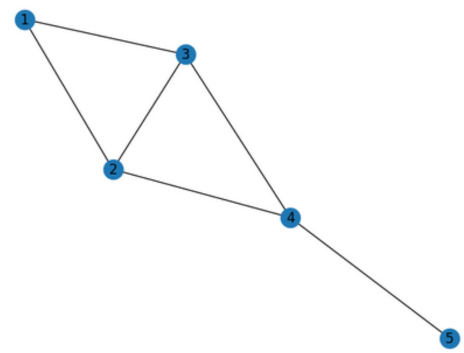
Step 1: Construct the Adjacency Matrix:
The adjacency matrix A for this network is:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Step 2: Compute the Degree Vector:
The degree vector d can be calculated by summing the elements in each row (or column) of the adjacency matrix.

$$d = [2\ 3\ 3\ 3\ 1]$$

Figure 1: Corresponding graph



Node 1 has a degree of 2 (connected to nodes 2 and 3).

Node 2 has a degree of 3 (connected to nodes 1, 3, and 4).

Node 3 has a degree of 3 (connected to nodes 1, 2, and 4).

Node 4 has a degree of 3 (connected to nodes 2, 3, and 5).

Node 5 has a degree of 1 (connected to node 4).

Step 3: Normalize the Degree Vector:
If we want to obtain the normalized degree centrality, we divide each element of the degree vector by N-1, where N is the total number of nodes in the network. In this case, N = 5.

$$c = [2/4\ 3/4\ 3/4\ 3/4\ 1/4] = [0.5\ 0.75\ 0.75\ 0.75\ 0.25]$$

Node 1 has a normalized degree centrality of 0.5.

Node 2 has a normalized degree centrality of 0.75.

Node 3 has a normalized degree centrality of 0.75.

Node 4 has a normalized degree centrality of 0.75.

Node 5 has a normalized degree centrality of 0.25.

In summary, the degree centrality for each node i is given by $d[i]$, and the normalized degree centrality is given by $c[i]$.

The advantage of using linear algebra for computing degree centrality is that it allows for efficient and scalable calculations, especially for large networks. Matrix operations can be performed efficiently using various numerical libraries or linear algebra packages.

For Directed Network

In directed networks, the computation of degree centrality is slightly different. In this case, there are separate measures for indegree centrality (number of incoming edges) and outdegree centrality (number of outgoing edges). The adjacency matrix may not be symmetric, and the degree vector would be calculated differently based on the directionality of edges.

In a directed, unweighted network, the indegree and outdegree of a node are calculated based on the number of incoming and outgoing edges, respectively. Let's define the adjacency matrix A for a directed network, where A[i][j] = 1 if there is a directed edge from node i to node j, and A[i][j] = 0 otherwise.

The indegree of a node i, denoted as indeg(i), is the number of incoming edges to that node. It is calculated by summing the elements in the i-th column of the adjacency matrix A. Mathematically,$indeg(i) = \Sigma A[j][i]$ for j = 1 to N, where N is the total number of nodes in the network.

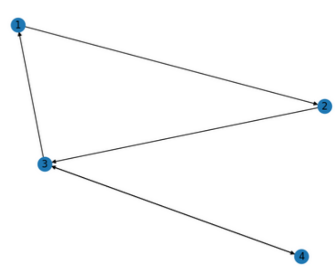
The outdegree of a node i, denoted as outdeg(i), is the number of outgoing edges from that node. It is calculated by summing the elements in the i-th row of the adjacency matrix A. Mathematically,

$$outdeg(i) = \Sigma A[i][j] for j = 1 to N.$$

For example, consider the following directed, network with 4 nodes and the corresponding adjacency matrix A:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Corresponding Graph :



To calculate the indegree and outdegree for each node:

Node 1:

indeg(1) = A[3][1] = 1

outdeg(1) = A[1][2] = 1

Node 2:

indeg(2) = A[1][2] = 1

outdeg(2) = A[2][3] = 1

Node 3:

indeg(3) = A[2][3] + A[4][3] = 2

outdeg(3) = A[3][1] + A[3][4] = 2

Node 4:

indeg(4) = A[3][4] = 1

outdeg(4) = A[4][3] = 1

So, in this example, the indegree and outdegree for each node are:

Node 1: indeg(1) = 1, outdeg(1) = 1

Node 2: indeg(2) = 1, outdeg(2) = 1

Node 3: indeg(3) = 2, outdeg(3) = 2

Node 4: indeg(4) = 1, outdeg(4) = 1

These calculations provide information about the number of incoming and outgoing connections for each node in the directed, unweighted network.

Applications of Degree Centrality:

1. **Identifying Influential Nodes:** Nodes with high degree centrality are often considered influential within a network. In social networks, individuals with a high number of connections may have more influence over information flow or decision-making processes. By identifying nodes with high degree centrality, researchers can focus their attention on influential entities.

2. **Understanding Network Structure:** Degree centrality provides insights into the overall structure of a network. Nodes with high degree centrality tend to be more connected and may serve as connectors or bridges between different parts of the network. This information is valuable for understanding the flow of information, identifying communities, or assessing the vulnerability of a network to disruptions.

3. **Targeted Interventions:** Degree centrality can help identify nodes that, if removed or targeted, would have the most significant impact on the network's connectivity. In fields like epidemiology or cybersecurity, targeting highly central nodes can be an effective strategy for preventing the spread of diseases or controlling the propagation of threats.

4. **Recommender Systems:** Degree centrality can be used in recommender systems to suggest items or connections based on the popularity or connectivity of nodes. For example, in a social network, suggesting friends with high degree centrality can lead to recommendations that are more likely to be accepted and valued by users.

Note: while degree centrality provides valuable information about the importance of nodes within a network, it has certain limitations. For instance, it fails to capture the quality or strength of connections, and it may not be suitable for all types of networks. Therefore, it is often used in conjunction with other centrality measures and network analysis techniques to gain a more comprehensive understanding of network dynamics.

2.2 BETWEENESS CENTRALITY:

Betweenness Centrality: Based on the network analysis notion of "betweenness centrality," a node's significance or centrality in the network is determined by its capacity to serve as a link or intermediary between other nodes. It measures how closely a node is located on the network's shortest pathways between adjacent pairs of nodes.

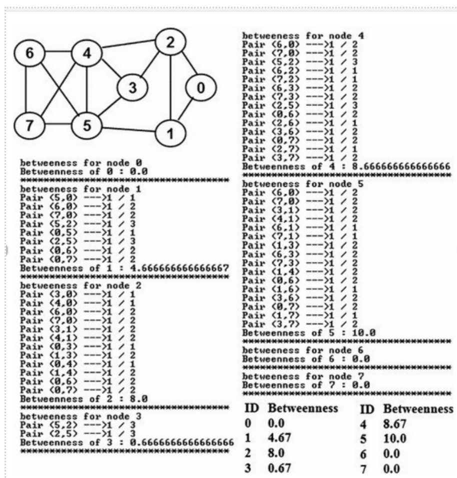
Computation of Betweenness Centrality: The computation of betweenness centrality comprises the calculating the shortest path that goes through each node in network. Calculating the fraction of shortest path.

1. INITIALIZATION:
   - We first Set each node's betweenness centrality value to zero.
2. PERFORM BFS:
   - To explore the network, launch a breadth-first search (BFS) from that node.
   - Now we find the shortest path to all other nodes and we keep track the number of shortest paths we encountered so far. That is from source node to destination node
3. STORING THE DISTANCES:
   - Store the distance from the source node and the total number of shortest routes to each node encountered throughout the BFS.
4. BACKTRACKING:

- Retrace the distances and the number of shortest routes from the destination nodes to the source node.
- The percentage of shortest routes that pass through each node along the path should be same.

5. OBTAINING THE FINAL VALUES:

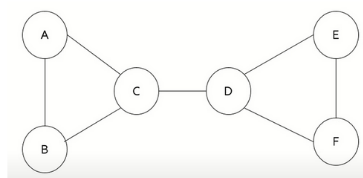
   - To determine the final betweenness centrality values, add up the obtained fractions for each node.
   - Now, divide the results by the total number of pairs of nodes (apart from the source node).
   - The betweenness centrality is given by the following expression.
   - $$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$
   - $\sigma_{st}(v))$ - Number of shortest paths from s to t passing through v
   - σ_{st} - Total number of shortest paths from s to t



Example to illustrate the calculation of betweenness centrality

The above example gives a clear notion to calculate the betweenness centrality. However, we get to infer that node 5 is much more localized than any other nodes.

here is another example



	σ_{uw}	$\sigma_{uw}(v)$	$\sigma_{uw}(v) / \sigma_{uw}$
(A,B)	1	0	0
(A,D)	1	1	1
(A,E)	1	1	1
(A,F)	1	1	1
(B,D)	1	1	1
(B,E)	1	1	1
(B,F)	1	1	1

	σ_{uw}	$\sigma_{uw}(v)$	$\sigma_{uw}(v) / \sigma_{uw}$
(D,E)	1	0	0
(D,F)	1	0	0
(E,F)	1	0	0

	σ_{uw}	$\sigma_{uw}(v)$	$\sigma_{uw}(v) / \sigma_{uw}$
(A,B)	1	0	0
(A,D)	1	1	1
(A,E)	1	1	1
(A,F)	1	1	1
(B,D)	1	1	1
(B,E)	1	1	1
(B,F)	1	1	1
(D,E)	1	0	0
(D,F)	1	0	0
(E,F)	1	0	0

Betweenness centrality of $C = 6$.

Applications of Betwenness Centrality

1. **Social Network Analysis:** In a social network, betweenness centrality can be used to identify individuals who act as information brokers or connectors. These individuals often have a high betweenness centrality because they serve as bridges between different groups or communities within the network. By identifying such individuals,

social network analysis can reveal key players who control the flow of information or influence in the network.

2. **Transportation Networks:**In transportation networks, betweenness centrality can help identify critical nodes or intersections that play a vital role in facilitating the movement of traffic. Nodes with high betweenness centrality are often located at strategic points where multiple routes converge. By focusing on these critical nodes, transportation planners can prioritize infrastructure improvements or allocate resources more effectively to ensure smooth traffic flow.

3. **Information Flow in Online Networks** In online networks, betweenness centrality can be used to analyze the spread of information or influence. Nodes with high betweenness centrality are likely to play a crucial role in disseminating information across the network. By targeting these influential nodes, marketers or advertisers can optimize their strategies to maximize the reach and impact of their messages.

4. **Power Grid Analysis** In the context of a power grid, betweenness centrality can identify key substations or nodes that act as critical pathways for the transmission of electricity. Nodes with high betweenness centrality are important for maintaining the overall stability and reliability of the power grid. Identifying these critical nodes helps in designing robust power grid systems and planning for contingencies.

2.3 Closeness Centrality

Closeness centrality is a measure that quantifies how close a node is to all other nodes in a network. It captures the concept of how quickly or efficiently information can spread from a particular node to other nodes in the network. Nodes with high closeness centrality are considered to have greater influence and prominence within the network due to their ability to access information or resources efficiently.

The closeness centrality of a node is calculated as the inverse of the average shortest path length from that node to all other nodes in the network. The formula for closeness centrality (C_{close}) of a node v in an undirected graph with N nodes is:

$$C_{\text{close}}(v) = \frac{N-1}{\sum \text{shortest\_path\_length}(v,u)}$$

Here, $\text{shortest\_path\_length}(v,u)$ represents the length of the shortest path between node v and node u, and $\sum$ denotes the sum of the shortest path lengths from node v to all other nodes in the network.

Step 1: Converting the data to a graph

This step involves representing the data as a graph using the adjacency matrix representation.

Step 2: Compute the shortest path from each vertex to every other vertex

Step 3: Calculate the closeness centrality of each vertex using the formula

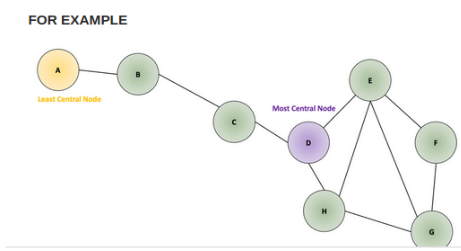
The closeness centrality score for each one of the nodes in the network is estimated as the reciprocal of the sum of the path of all other nodes.

$$C(x) = \frac{1}{\Sigma_y d(y, x)}$$

Step 4: Normalize the closeness centrality to account for the size of the network Normalize accordingly.

$$closeness\ centrality\ score(u) = \frac{number\ of\ nodes - 1}{\Sigma(distance\ from\ u\ to\ all\ other\ nodes)}$$

Example



	A	B	C	D	E	F	G	H	Normalized Score
A	0	1	2	3	4	5	5	4	0.35
B	1	0	1	2	3	4	4	3	0.39
C	2	1	0	1	2	3	3	2	0.50
D	3	2	1	0	1	2	2	1	**0.58**
E	4	3	2	1	0	1	1	1	0.54
F	5	4	3	2	1	0	1	2	0.39
G	5	4	3	2	1	1	0	1	0.41
H	4	3	2	1	1	2	1	0	0.50

Directed Network

In the case of directed networks, most software only calculates the outbound closeness centrality, considering the outbound relationships: the most important node is the one that can reach all other nodes in the network most quickly and helps define good broadcasters (e.g., Instagram), which is important in diffusion processes.

The inbound closeness centrality accounts for the inbound relationships and can be useful if we want to find the optimal location for a customer or improve the search results of a website; in these cases, the incoming parts are meaningful.

Website searches represent another good closeness centrality example to better grasp the difference between the outbound and inbound closeness centrality score: we can have a high closeness centrality from outgoing website links but low closeness centrality from incoming website links.

Weighted Network

We can also calculate closeness centrality in weighted networks, where the distances are not always going to be the same. In this case, the shortest path is not the one with the least hops but the one with the minimum total weighted distance. The normalized closeness centrality score for node B is

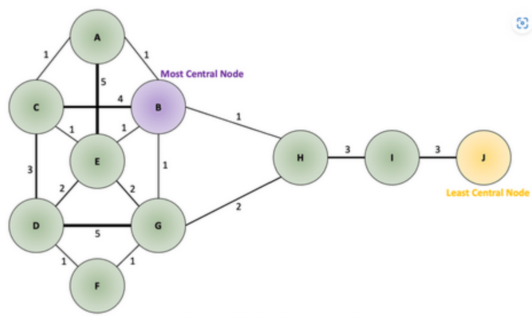


Figure 2: Weighted Graph Network Example.

	A	B	C	D	E	F	G	H	I	J	Normalized Score
A	0	1	1	4	2	3	2	2	5	8	0.32
B	1	0	2	3	1	2	1	1	4	7	**0.41**
C	1	2	0	3	1	4	3	3	6	9	0.28
D	4	3	3	0	2	1	2	4	7	10	0.25
E	2	1	1	2	0	3	2	4	7	10	0.28
F	3	2	4	1	3	0	1	3	6	9	0.28
G	2	1	3	2	2	1	0	2	5	8	0.35
H	2	1	3	4	4	3	2	0	3	6	0.32
I	5	4	6	7	7	6	5	3	0	3	0.19
J	8	7	9	10	10	9	8	6	3	0	0.13

$$\frac{9}{1+2+3+1+2+1+1+4+7} = 0.41$$

, which turns out to be the most central node in this network, and for node

$$J\text{, it is } \frac{9}{8+7+9+10+10+9+8+6+3} = 0.13$$

, resulting in being the farthest node.

Applications of Closeness Centrality

Infrastructure and Communication Networks: Closeness centrality is used in analyzing and optimizing various types of networks, including power grids, telecommunication networks, and internet routing. Identifying nodes with high closeness centrality helps in improving network efficiency, reducing communication delays, and enhancing the overall robustness of the network.

Urban Planning: Closeness centrality can be applied in urban planning to identify key locations for services and facilities. For example, it can assist in determining the optimal placement of schools, hospitals, or public transportation stations to ensure that these amenities are easily accessible to the majority of the population.

Influence Maximization: Closeness centrality is employed in influence maximization problems, where the goal is to identify a set of influential nodes that can maximize the spread of information, ideas, or influence in a social network. Nodes with high closeness centrality are often targeted as potential seed nodes to initiate cascades of influence or diffusion processes.

Financial Networks: Closeness centrality can be applied in analyzing financial networks, such as interbank networks or stock market networks. Nodes with high closeness centrality in these networks may represent critical financial institutions or stocks that have a significant impact on the overall stability and functioning of the financial system.

Limitations of Closeness Centrality Algorithm

- The position of the node is related to the entire network (big advantage!).
- Closeness Centrality is sensitive to changes in the network.
- All nodes must be reachable as a constraint to estimate the closeness centrality score.

2.4 Eigenvector Centrality

A measure of a node's significance or centrality in a network is called eigen centrality, also known as eigenvector centrality. It measures a node's prominence or impact based on connections to other nodes that are closely related to it. A node is seen as significant if it is linked to other important nodes, which is the basis for the idea of eigen centrality.
In a network, nodes represent entities (such as individuals, websites, or organizations), and edges represent the connections or relationships between these entities. Eigenvector centrality assigns a numerical value to each node, indicating its relative importance based on both its direct connections and the importance of its neighbouring nodes.
The calculation of eigenvector centrality involves iteratively determining the centrality values of nodes in a network until convergence is achieved. The underlying assumption is that the centrality of a node is proportional to the sum of the centrality scores of its neighbours.

Computation of Eigenvector Centrality

1. **Adjacency Matrix:** The first step is to construct an adjacency matrix A, which represents the connections between nodes in the network. This matrix is typically square, with dimensions equal to the number of nodes in the network.

2. **Eigenvalues and Eigenvectors:** An eigenvector v of a square matrix A is a non-zero vector that satisfies the equation:

$$A \times v = \lambda \times v$$

where λ is a scalar known as the eigenvalue corresponding to the eigenvector v. In the context of eigenvector centrality, we are interested in finding the principal eigenvector, which corresponds to the largest eigenvalue.

3. **Power Iteration Method:**

   The power iteration method is commonly used to find the principal eigenvector and eigenvalue. It involves iteratively multiplying the adjacency matrix A by a vector until convergence is achieved.

   Let's denote the current estimate of the eigenvector as x_0, an initial guess that can be arbitrary or uniform. We can then iteratively update x_k using the following steps:

   - Multiply the current estimate x_k by the adjacency matrix A: $y_k = A \cdot x_k$.
   - Normalize the resulting vector y_k to maintain its magnitude: $x_{k+1} = \frac{y_k}{\|y_k\|}$, where $\|y_k\|$ represents the Euclidean norm of y_k.
   - Repeat these steps until convergence is reached. Convergence occurs when the difference between x_k and x_{k+1} falls below a predefined threshold.

   After convergence, the resulting vector x will be the principal eigenvector corresponding to the largest eigenvalue of the adjacency matrix A. It represents the eigenvector centrality scores of the nodes in the network.

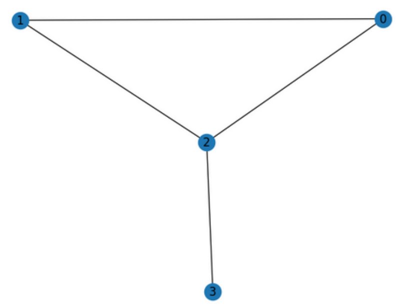
4. **Normalization:** To ensure meaningful comparisons between nodes, the resulting eigenvector x should be normalized. This is done by dividing each element of x by the sum of all its elements.

5. **Interpreting Eigenvector Centrality:** The values in the normalized eigenvector x represent the eigenvector centrality scores of the nodes in the network. Nodes with higher scores are considered more central or influential within the network.

For Example
Suppose we have the following network represented by its adjacency matrix:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Corresponding Graph :



1. **Constructing the Adjacency Matrix:** We start by representing the network as an adjacency matrix. The matrix A above indicates the connections between the nodes. For example, there is a connection between nodes 0 and 1, so A[0][1] and A[1][0] are set to 1.

2. **Initial Centrality Assignment:** We assign an initial centrality score to each node. Let's start with uniform values, so each node has an initial centrality score of 1. We can represent it as a list:

$$x_0 = [1, 1, 1, 1]$$

3. **Computing Centrality Scores:** We compute the centrality score for each node based on the centrality scores of its neighbours.

$$y_0 = A \times x0$$

Let's go through the calculations:

For node 0:

$$y_0[1] = (0 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) = 2$$

For node 1:

$$y_0[2] = (1 \times 1) + (0 \times 1) + (1 \times 1) + (0 \times 1) = 2$$

For node 2:

$$y_0[3] = (1 \times 1) + (1 \times 1) + (0 \times 1) + (1 \times 1) = 3$$

For node 3:

$$y_0[4] = (0 \times 1) + (0 \times 1) + (1 \times 1) + (0 \times 1) = 1$$

The updated centrality scores after the first iteration become:

$$y_0 = [2, 2, 3, 1]$$

4. **Normalization:** We normalize the centrality scores by dividing each score by dividing the Euclidean norm of y_0 This step ensures that the centrality values lie between 0 and 1, facilitating interpretation and comparison.
   The Euclidean norm of y_0 is 4.242.
   Thus, the normalized centrality scores become:

   $$x_1 = [2/4.242,\ 2/4.242,\ 3/4.242,\ 1/4.242] = [0.47140452 \quad 0.47140452 \quad 0.70710678 \quad 0.23570226]$$

5. **Convergence Check:**

   We check whether the centrality scores have converged. Convergence occurs when the difference between the centrality scores of consecutive iterations falls below a predefined threshold.

   For checking convergence, We would then compare these updated scores with the previous scores and check if the difference falls below the threshold. If it does, we consider the centrality scores to have converged.

   For the above example

   Iteration 1

   $$y_0 = [2 \quad 2 \quad 3 \quad 1]$$

   Eucledian Norm of y_0 is 4.242640687119285

   $$x_1 = [0.47140452 \quad 0.47140452 \quad 0.70710678 \quad 0.23570226]$$

   Iteration 2

   $$y_1 = [1.1785113 \quad 1.1785113 \quad 1.1785113 \quad 0.70710678]$$

   Eucledian Norm of y_1 is 2.160246899469287

   $$x_2 = [0.54554473 \quad 0.54554473 \quad 0.54554473 \quad 0.32732684]$$

   Iteration 3

   $$y_2 = [1.09108945 \quad 1.09108945 \quad 1.41841629 \quad 0.54554473]$$

   Eucledian Norm of y_2 is 2.165750722146062

   $$x_3 = [0.50379272 \quad 0.50379272 \quad 0.65493054 \quad 0.25189636]$$

   Iteration 4

   $$y_3 = [1.15872326 \quad 1.15872326 \quad 1.2594818 \quad 0.65493054]$$

Eucledian Norm of y_3 is 2.1680654081953334

$$x_4 = [0.53445032 \quad 0.53445032 \quad 0.58092427 \quad 0.30208062]$$

Iteration 5

$$y_4 = [1.11537459 \quad 1.11537459 \quad 1.37098127 \quad 0.58092427]$$

Eucledian Norm of y_4 is 2.169143514227027

$$x_5 = [0.51420046 \quad 0.51420046 \quad 0.63203806 \quad 0.26781274]$$

Iteration 6

$$y_5 = [1.14623852 \quad 1.14623852 \quad 1.29621366 \quad 0.63203806]$$

Eucledian Norm of y_5 is 2.169646850220248

$$x_6 = [0.52830649 \quad 0.52830649 \quad 0.59743071 \quad 0.29130919]$$

Iteration 7

$$y_6 = [1.1257372 \quad 1.1257372 \quad 1.34792218 \quad 0.59743071]$$

Eucledian Norm of y_6 is 2.1698815988657194

$$x_7 = [0.51880121 \quad 0.51880121 \quad 0.62119619 \quad 0.27532871]$$

Iteration 8

$$y_7 = [1.1399974 \quad 1.1399974 \quad 1.31293114 \quad 0.62119619]$$

Eucledian Norm of y_7 is 2.169991018795449

$$x_8 = [0.5253466 \quad 0.5253466 \quad 0.60503989 \quad 0.28626671]$$

Iteration 9

$$y_8 = [1.13038649 \quad 1.13038649 \quad 1.33695991 \quad 0.60503989]$$

Eucledian Norm of y_8 is 2.1700420071005753

$$x_9 = [0.52090535 \quad 0.52090535 \quad 0.61609863 \quad 0.27881483]$$

Iteration 10

$$y_9 = [1.13700398 \quad 1.13700398 \quad 1.32062553 \quad 0.61609863]$$

Eucledian Norm of y_9 is 2.1700657639635854

$$x_{10} = [0.52394909 \quad 0.52394909 \quad 0.60856475 \quad 0.28390781]$$

Iteration 11

$$y_{10} = [1.13251384 \quad 1.13251384 \quad 1.33180599 \quad 0.60856475]$$

Eucledian Norm of y_{10} is 2.1700768322831054

$$x_{11} = [0.5218773 \quad 0.5218773 \quad 0.61371375 \quad 0.28043466]$$

Iteration 12

$$y_{11} = [1.13559105 \quad 1.13559105 \quad 1.32418926 \quad 0.61371375]$$

Eucledian Norm of y_{11} is 2.1700819888683838

$$x_{12} = [0.52329408 \quad 0.52329408 \quad 0.61020241 \quad 0.28280671]$$

Iteration 13

$$y_{12} = [1.13349649 \quad 1.13349649 \quad 1.32939487 \quad 0.61020241]$$

Eucledian Norm of y_{12} is 2.1700843912228844

$$x_{13} = [0.5223283 \quad 0.5223283 \quad 0.61260054 \quad 0.28118833]$$

Iteration 14

$$y_{13} = [1.13492884 \quad 1.13492884 \quad 1.32584493 \quad 0.61260054]$$

Eucledian Norm of y_{13} is 2.1700855104271644

$$x_{14} = [0.52298807 \quad 0.52298807 \quad 0.61096437 \quad 0.28229327]$$

Iteration 15

$$y_{14} = [1.13395244 \quad 1.13395244 \quad 1.32826942 \quad 0.61096437]$$

Eucledian Norm of y_{14} is 2.1700860318384283

$$x_{15} = [0.52253801 \quad 0.52253801 \quad 0.61208145 \quad 0.28153924]$$

we can see that the values are almost same after iteration 11.
So the final centrality vector will be x_{11}

6. **Final Centrality Values:** The resulting centrality scores represent the eigenvector centrality of each node in the network. Nodes with higher centrality scores are considered more important or influential within the network.
   In this example, the final eigenvector centrality scores are:

$$x_{11} = [0.5218773 \quad 0.5218773 \quad 0.61371375 \quad 0.28043466]$$

   These scores indicate the relative importance of each node within the network. Node 3 has the highest centrality score (0.6137), indicating that it is the most central or influential node in the network. Nodes 1 and 2 have similar centrality scores (0.5218), suggesting a similar level of centrality, while node 4 has the lowest score (0.2804) indicating a lower level of centrality or influence.

Physical Significance of Eigenvector Centrality

Eigenvector centrality is a measure used in network analysis to quantify the importance or influence of a node within a network. The physical significance of eigenvector centrality can vary depending on the specific context in which it is applied. Here are a few potential interpretations:

1. **Influence or Prestige:** In certain networks, eigenvector centrality can be interpreted as a measure of influence or prestige. Nodes with high eigenvector centrality are considered more influential or prestigious because they are connected to other important or influential nodes. For example, in a social network, a person with high eigenvector centrality may be seen as influential because they are connected to other well-connected individuals.

2. **Information Flow:** Eigenvector centrality can also be related to the flow of information or resources in a network. Nodes with high eigenvector centrality can act as central points for information dissemination, where information or resources tend to flow through these nodes more easily. This interpretation is particularly relevant in communication networks, transportation networks, or any system where the efficient flow of information or resources is essential.

3. **Structural Importance:** Eigenvector centrality can reflect the structural importance of a node within a network. Nodes with high eigenvector centrality tend to occupy critical positions within the network, often serving as bridges or connectors between different groups or clusters of nodes. Removing or targeting nodes with high eigenvector centrality can have a significant impact on the overall connectivity and integrity of the network.

4. **Power or Control:** In some cases, eigenvector centrality can be associated with power or control within a network. Nodes with high eigenvector centrality may have the ability to control or influence the behaviour of other nodes in the network. This interpretation is often used in social

or organizational networks, where individuals with high eigenvector centrality may have more control over decision-making processes or access to valuable resources.

Application of Centrality Metrics in Real-Life Scenarios

Centrality metrics have been studied since the 1940s. Even in the late 1970s, there exists a rich volume of studies discussing and experimenting with centrality. Below we have elaborated on how centrality has been studied in various fields.

Chemistry

Chemical process plants can be represented by networks in which centrality metrics are used to identify more important units and controllers.

Anthropology

Network centrality was first investigated in Anthropology by studying human behaviors in groups. Many group-based decision-making research communities have studied centrality metrics to measure influence and/or power of a group or organization. In the recent Anthropology research, Collins and Durington discussed 'networked anthropology' by using diverse multimedia and OSN platforms. In addition, how community centrality affects scholarly activities in social science has been studied in Anthropology.

Geography

Historical geographers were interested in how the centrality of a region (e.g., Moscow) can affect dominance and evolution of the region in which the area can be described based on graph theory. They studied urban street networks based on graph theory in order to identify important areas in terms of the influence of topology and geo-referenced data extracted from the network.

Economics

Souma et al. studied business networks to investigate the probability of business networks becoming scale-free and the effect of the merger among banks on the cliquishness of companies or the separation between two companies. Mayer also investigated how social and economic factors (e.g., economic incentives or socioeconomic background) can introduce the changes in social network structure and its composition which were measured by centrality metrics (e.g., Eigenvector centrality).

Psychology

Centrality metrics have been used to measure socio-cognitive aspects of human behavior in various contexts. Kameda et al. defined a person's power in a group based on his/her centrality measured by the degree of information the person shares with others. The person's influence based on network centrality has been shown to be critical to forming consensus in the decision-making process. Lee et al. looked at how a person's centrality in a network position affects consumer influence as well as susceptibility to the influence of others.

Biology

Centrality metrics have been used in Biology in selecting central nodes, such as pathogen-interacting, cancer, ageing, HIV-1, or disease-related or immune-related proteins in gene regulatory networks, protein-interaction networks, and metabolic networks.

Management

Centrality metrics have been investigated to identify the key factors to be successful in business management. The management research has investigated how a founder's centrality affects the top management group, the group's culture and vision, and how network centrality is critical to increasing financial performance.

Computer Science

Centrality metrics have been highly leveraged and investigated for diverse applications in the computer science domain. For example, centrality metrics are used in mobile social network applications, visual reasoning in online social networks, water network distribution, or traffic management for space satellite networks.

Psychiatry

Network science has been applied in Psychiatry under the name of Network Psychiatry based on computational models to investigate the structure of psychiatric disorders which are treated as complex systems. Centrality metrics have been considered to measure 'functional connectivity' in a brain connectome. They investigated the relationship between the extent of centrality and certain diseases or body conditions/characteristics. Their findings backed up how the centrality in the brain connectome can be used as the underlying physiological mechanisms to study 'neurodegenerative and psychiatric disorders.' Fried et al. also used centrality metrics to determine the centrality of the Diagnostic and Statistical Manual of Mental Disorders (DSM) symptoms and non-DSM symptoms where a network consists of 28 depression symptoms. In this work,

centrality is used as an indicator of the relationships between different depression symptoms.
Let us see some particular and latest implementations of Centrality Networks.

Application 1: Marketing for a product- pay those who are most popular

In the simplest case, the number of a network member's direct contacts is a useful indicator of centrality. The advantage of this interpretation of an actor's centrality, with degree centrality (DC) as its standard representative is the fact that the results are relatively easy to interpret and communicate.

A second approach is based on the idea that nodes that have a short distance to other nodes and consequently are able to disseminate information on the network very effectively, take a central position in the network. A representative of this approach is closeness centrality (CC), where a person is seen as centrally involved in the network if he requires only few intermediaries for contacting others and thus is structurally relatively independent. Accordingly, the calculation of this CM includes the length of the shortest paths to all other actors in the network. Further developments of CC even use the length of all paths between the actors for the calculation.

A third approach, however, equates centrality with the control of the information flow which a member of the network may exert, based on his position in the network. Here, it is assumed implicitly that the communication and interaction between two not directly related actors depends on the intervening actors. The most prominent representative of this concept is betweenness centrality (BC), where the determination of an actor's centrality is based on the quotient of the number of all shortest paths between actors in the network that include the regarded actor and the number of all shortest paths in the network.

The common characteristic of all networking concepts presented so far is that only little or no attention is paid to indirect contacts, meaning they are not or only indirectly included in the quantification of an actor's centrality. This is where the so-called influence measures come into play. These CM consider actors to be centrally involved in the network if their directly connected network members are in relationship with many other well-connected actors. Some of the best known of these recursively defined CM are the eigenvector centrality (EC).

Besides these representatives of the four basic concepts of centrality, a plethora of other CM has been defined over the years enable the integration of edge weights or of directional connections or are suitable for specific applications and network types.

For the mathematical calculation of each CM, different algorithms have been developed which may vary significantly in terms of complexity. While the DC only requires to count the direct contacts of the n nodes in the network (complexity of O(n)), the complexity of BC in unweighted graphs amounts to O(nm), where m is the number of edges in the network.

Application 2: Finding which scientific standard is most connected to others

A general problem associated with the scientific standard analytics is that this particular standards committee has a large number of standards and is losing track of how they all relate to each other. It is a basic configuration management problem, such as the impact of change. It is generally acknowledged that visualisations have some benefits when it comes to making sense of large and complex non-visual datasets.

To improve the situation, the scientific standards management section at the University of Technology Sydney (UTS) has initiated a joint project with the authors to investigate how graph (or network) visualisation and graph centrality metrics can be used to help analysts make sense of scientific standards data sets quickly and accurately.

Standard Relations	Edge Weight
Normative Standard—Normative Standard	3
Normative Standard—Informative Standard	2
Informative Standard—Informative Standard	1

In this study, raw data attributes such as the standard name and related reference name are kept as vertices. Edges have directions and are represented depending on their 'reference' relations. Edge weights are calculated based on the standard/reference types defined in Table 2. Edges between normative standards are the most important connections as normative standards contain a number of compulsory requirements for other standards. Therefore, the weight of an edge between two normative standards is defined as 3. For example, the weight of the edge that connects standard 12207:2008 and 9126:1991 is 2, since one standard is normative and the other one is informative.

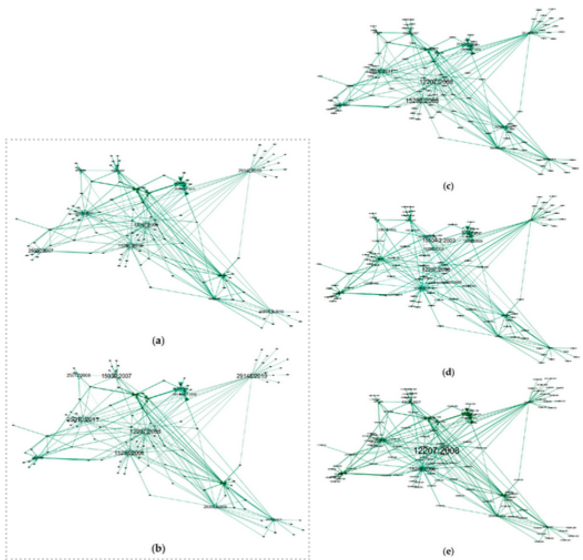
In this study, a standard dataset is a labelled directed weighted graph G = (V, E, w), where V is the set of nodes, E is the set of edges and w is the weight function. Six centrality metrics are adopted in experiments to examine the standard's network.

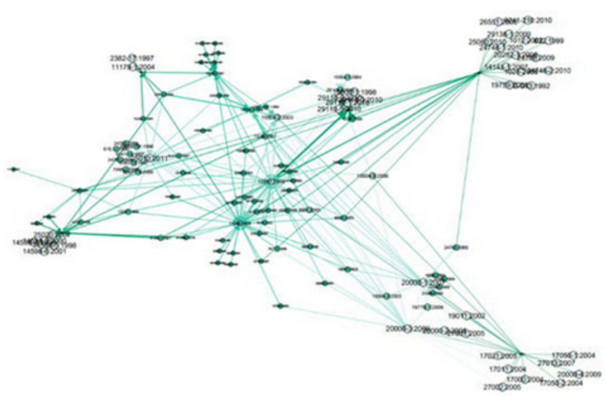
The Eigenvector centrality concept is adopted as a ranking measure to analyse the importance of standards.

Eventually, relevant graph models are generated, and

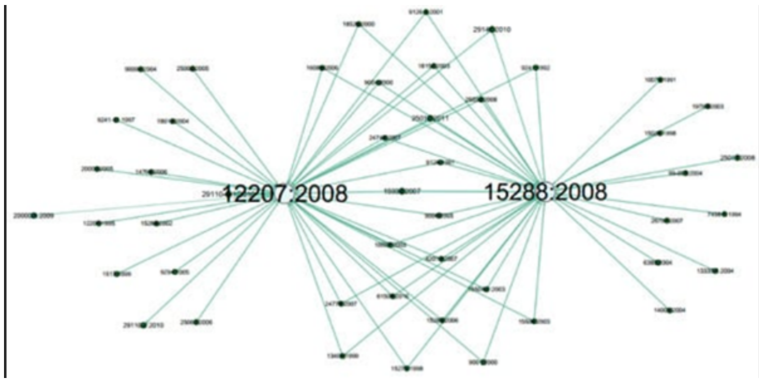
$$G1 = (V1, E1)(|V1| = 121, |E1| = 248)$$

represents the entire network of 121 scientific standards collected.





Visualisations of a standard graph for all measurements (FA2, —E— = 121, —V— = 248). Node size, colour depth and label size indicate standard rankings. For example, nodes with the larger size, darker green colour and larger label size are more important standards with higher rankings. (a) PageRank measurement (25030:2007 is the highest-ranking standard); (b) Eigenvector centrality measurement (25010:2011 is the highest-ranking standard); (c) Weighted In-Degree measurement (12207:2008 is the highest-ranking standard); (d) Weighted Out-Degree measurement (12207:2008 is the highest-ranking standard); (e) Betweenness measurement (12207:2008 is the highest-ranking standard); (f) Closeness measurement (Top ten ranking standards have the same value).



Six visualisations that are generated for the six centrality metrics are shown. As can be seen from the gifure, minor differences are among all these layouts except the Weighted Out-Degree and Closeness measurements. There is nearly 80*percent* match in the PageRank, Eigenvector, Weighted In-Degree and Betweenness centrality metrics, while the remaining two metrics represent different results.

From below figure, it can be seen that the standard which has the strongest connection to 12207:2008 is 15288:2008, followed by 29110-4-1:2010. Also 12207:2008 has the highest correlation with 15288:2008, while 15939:2007 has the second highest correlation with it.

Application 3: Criminal Investigation

Graph databases are revolutionizing criminal activity analysis. Some crimes happen on a small, opportunistic scale. But the kind of crime that police work together to track and take down tends to happen on a large scale with many interconnected people, gangs, businesses, and even locations—which means that it doesn't tend to happen in silos. The graph solution Putting data into graphs provides a natural and efficient way of identifying criminal networks and looking for patterns. By applying graph based algorithms like centrality, it becomes easier to look for vulnerable people in the graph, discover more insights regarding locations, and even look for important people and potential criminal gangs. For example, by applying betweenness centrality, users can find the "weakest link," meaning the vertex that the graph relies upon. If you remove that vertex, the entire graph may fall apart—meaning you may have just found the linchpin of a criminal gang.

Application 4: Contact Tracing

Disease contact tracing is a critical activity worldwide. People become ill with highly infectious, new diseases and continue to live their normal lives—spreading that disease everywhere they go. When someone is diagnosed, the race to find everyone who has been in contact with that sick person to ask them to quarantine, becomes a race against time. Contact tracers must be able to do their work as quickly as possible to stop further spread of the disease.

The graph solution Graph databases, with their heavy emphasis on relationships, are ideal to use for analyzing disease patterns. Analysts can input information on the people who have tested ill, the family members and friends they've interacted with, and the places they've visited, to rapidly locate hotspots and connections. In this way, analysts can work more quickly to isolate those who are ill and prevent further disease outbreaks.

There are three levels to contact tracing with graphs.

First, there is the need to understand people's relationships, communities, and the places they visit which graphs can make clear if provided with enough mobile data.

Secondly, graphs must find the possible spread—which means looking at potential links between people who might spread the disease. Did the person travel by bus? Can we identify everyone on the bus?

Third, contact tracers must find "super spreaders" and rush to isolate those people first. This involves finding the people who have wide and dense contacts and who are likely to have links to many different communities. This involves exploring the graphs with notions of centrality and betweenness, to find the highly connected people.

2.5 COMPARATIVE ANALYSIS:

Centrality Metric	**Category**	**Meaning**	**Complexity**
Degree	Local Centrality Metrics	Popularity	$O(n+m)$
Betweenness	Global Centrality	Measuring the influence of a node as a broker	$O(mn)$
Closeness	Global Centrality	Reciprocal of distance sum of a node to all other nodes	$O(mn)$
Eigenvector	Iterative Centrality Metrics	Importance of neighboring nodes determines node's importance	$O(n^3)$

3 State Of Art Literature (SOTA) :

Centrality metrics have been employed in many projects in the real-world scenario, however, the most prominent one is SNA - Social Network Analysis.We have demonstrated one model through our codes. Another citation is described in the research paper attached below.

3.1 Limitations

1. Node centrality determines influence in terms of connectivity, communicability, and controllability in each network. However, node connectivity is not commonly aligned with the capacity to deal with traffic (e.g., communicability) because nodes with high connectivity are often congested.

2. Although a large volume of centrality metrics has been developed so far, only common centrality metrics have been used, such as degree, betweenness, closeness, PageRank, which has been developed for several decades ago. Although degree is a simple metric, other metrics, such as betweenness or clustering coefficient, require high complexity with high running time.

3. Centrality metrics analysis relies on the fact that larger connectedness denotes more importance. However, this may not be the case in certain scenarios, for example, cyber security systems where smaller components might have important nodes that need to be protected on a priority basis compared to larger components.

3.2 Future Research Opportunities

1. **More Efficient Centrality Metrics are Required**

   According to our study of centrality metrics so far, the algorithms that have a nominal complexity provide limited data about the data like degree centrality based on which we cannot really make trust-worthy decisions. Relevant metrics like betweenness and closeness that are representative of a broader meaning of centrality have a high complexity that limits us from using them in larger networks.

2. **More Meaningful Metrics are Needed to Measure Network Resilience**

   The size of the giant component, as a common metric to measure network resilience, does not reflect a broader concept of network resilience. Network resilience is how adaptable a network is to deal with sudden changes or attacks/failures (i.e., adaptability), how tolerant the network is to prevent its failure against attacks or failures (i.e., fault tolerance), and how easily recoverable the network is from attacks or failures (i.e., recoverability). We need to develop metrics that can measure network resilience embracing adaptability, fault tolerance, recoverability, or other properties based on system requirements.

3. **Algorithms for Deciding on a Centrality Metric for Applications**

   We can analyze different centrality metrics to identify which metric would be more powerful under different network conditions. In addition, more comprehensive, diverse, larger, and real network topologies can be considered to obtain more meaningful findings and provide generalizable guidelines for selecting useful centrality metrics in each application.

4 REFERENCES

References for closeness centrality from gfg

refrences of closeness centrality

Refrences for Betwenness Centrality

this includes every centrality in detail

This is FaceBook dataset

17 Use Cases for Graph Databases and Graph Analytics (oracle.com)

(PDF) Eigenvector centrality and its application in research professionals' relationship network (researchgate.net)

Symmetry — Free Full-Text — Applying Graph Centrality Metrics in Visual Analytics of Scientific Standard Datasets (mdpi.com)

17 Use Cases for Graph Databases and Graph Analytics (oracle.com)

A Critical Review of Centrality Measures in Social Networks — SpringerLink