Pure Sciences International Journal of Kerbala

Analysis of Large Networks Using Statistical and Mathematical Techniques: An Applied Study Using MATLAB

Ali Mohamed Husein Nasralla*

College of Education for Pure Sciences, University of Kerbala, Karbala, Iraq

PAPER INFO

Paper history:
Received: 14 February 2025
Accepted: 24 February 2025
Published: 30 June 2025

Keywords:
Network analysis, degree centrality, median centrality, closeness centrality.

ABSTRACT

This paper presents an analytical study of networks using mathematical and statistical techniques, with a focus on practical applications using MATLAB. The study aims to understand the internal structure of networks and analyze the nodes with the highest influence using centrality factors such as degree centrality, median centrality, closeness centrality, the experience conducted with MATLAB-supported applications for network representation and extraction of statistical indicators.

1. INTRODUCTION

Networks have become a fundamental component of a wide range of systems. The growth and complexity of these networks has created an urgent need for advanced analytical methods to understand their structure and function.

Large-scale networks are characterized by numerous nodes and complex interconnections. Network analysis requires advanced mathematical and statistical techniques that enable the definition of essential components, the display of hidden structures, and the discovery of elements which may not be noticed directly.[10]

Network analysis incorporates the concept of node centrality, which reflects the importance or influence of each individual node within a network. Various centrality measures—such as degree centrality, closeness centrality, betweenness centrality, pagerank, and eigenvector centrality—offer diverse perspectives on a node's role in information flow, communication, and control within a system. These measures are active in fields such as computer science .

One of the key aspects of network analysis is community detection, which contains identifying sets of nodes that are more densely connected than the rest of the network. Algorithms such as the Louvain method are significantly used for big data due to their efficiency and scalability.

In this study, we present an overall approach to analyzing large networks by combining mathematical concepts with practical applications using the MATLAB.The MATLAB provides a powerful set of built-in functions that make it suitable for simulating, analyzing, and interpreting network structures. By using algorithms to data network's structures, this study aims to identify impact nodes for the underlying network.

2. PREVIOUS STUDIES

The developing of extensive networks has get significant academic interest across a common of fields. Some studies have used mathematical and statistical methods for evaluating the architecture and functionality of such networks, spatially in the discovering of influential nodes and the study of community structures.

Estrada and Rodríguez-Velázquez explained the notion of subgraph centrality, which matures the significance of a node based on its sharing in different subgraphs of the network, thereby

*Corresponding Author Institutional Email:
ali.nasralla@uokerbala.edu.iq (Ali Mohamed Husein Nasralla)

providing more deep insights than traditional centrality metrics [1]. Klein used a comparison analysis of conventional centrality indicators, including degree, closeness, and betweenness centrality, accentuating their mathematical underpinnings and applicability across various contexts [3].

Latora, Nicosia, and Russo presented a thorough comprehensive study on complex network theory, showing both foundational principles and practical applications such as centrality analysis and community detection, often supplemented by MATLAB code exemplifications [2]. Gómez developed the discourse by applying centrality analysis to business networks, confirming the workable importance of recognize axial nodes in real world data [4].

In the scope of wireless sensor networks, Mbiya, et al. proposed an active routing algorithm that authority centrality mensuration to promote communication pathlane. Their method showing practical efficacy through simulations executed using MATLAB [5].

An important progression in community detection was introduced by Blondel et al. through the Louvain algorithm, that activity reveals hierarchical modular structures within extensive networks, rendering it appropriate, for large data [6].

Newman provided an elaborate mathematical methods for understanding, various network attributes such as modularity and centrality [7]. Additionally, Wasserman et al. understood, the foundational essentials for social network analysis, presenting rigorous statistical and mathematical approaches for examining network [8].

These previous studies constitute the bedrock of research in network science and provide a robust foundation for the present researches, which builds upon these approaches employing applied tools in MATLAB for the analysis of overall and intricate networks.

3. RESEARCH METHODOLOGY

The research is divided into two main parts:

3.1 Theoretical part

This part focuses on the mathematical and statistical ideas used to analyze large networks. It covers the definition of a network and its basic elements, and describes the significance of centrality measures such as degree , closeness , betweenness centrality. It also labels detecting communities' techniques within networks, with a focuse on the benefit of the Louvain algorithm in analyzing big data.

This section puts the theoretical base for the future application of analytical tools using MATLAB.

The Louvain algorithm is a fast and efficient method for subdividing large networks into communities. It is based on enhancing the modularity measure, which measures the density of links within each community compared to the links among communities.

Node Degree is a concept used in network analysis (whether computer, social, or otherwise) to indicate the number of direct connections a node has with other nodes.[4]

1. Importance of Node Degree

- In computer networks:
  - Node degree reflects the number of devices or paths connected to a router or access point.
  - Nodes with a high degree may be critical points in the network.
- In social networks:
  - Node degree represents the number of connections (friends or followers) an individual has.

It is used to analyze influence or importance in a network.

- In transportation networks:

The degree of a node determines the number of roads or paths leading to a given point.

2. Types of Node Degree:

- Node degree in an undirected network:
  - The number of edges connected to the node.

For example: If a node is connected to three other nodes, its degree is 3.

- Node degree in a directed network:
- In-degree: The number of edges terminating at the node.
- Out-degree: The number of edges starting from the node.

Total degree = In-degree + Out-degree.

3. Equations for Calculating Node Degree

- Undirected Networks:

\text{Degree} = \text{Number of Edges Associated with the Node}

- Directed Networks:

\text{In-Degree} = \text{Number of Edges Entering the Node}

\text{Out-Degree} = \text{Number of Edges Outgoing from the Node}

4. Practical Example

Undirected Network:

```
A -- B -- C
     |
     D
```

Node Degree:
A: 1
B: 3
C: 1
D: 1

Directed Network:

```
A → B → C
  ↓     ↑
  D ←   E
```

In and Out Degree:
A: In=1, Out=1
B: In=1, Out=2
C: In=1, Out=0
D: In=1, Out=1
E: In=1, Out=1

Closeness Centrality

is a measure in network analysis used to determine the "closeness" of a node. It is based on the total distance between a given node and all other nodes in the network. Simply put, it measures how quickly a node can reach other nodes in the network.[5]

The closeness centrality of a given node is given by the following relationship:

$$c(v) = \frac{1}{\sum_{u \neq v} d(v,u)} \tag{1}$$

Where:

- d(v,u): is the shortest distance between nodes v and u.
- The sum is calculated for all other nodes u in the network.

Concept:

- Proximity to a node: If a node is close to most other nodes (i.e., the distances between it and other nodes are short), its centrality will be high.
- Far from other nodes: If a node is far from other nodes (i.e., the distances are long), its closeness centrality will be low.

The Importance of Closeness Centrality

- Social Networks:
  - Helps identify individuals who can spread information or influence others most quickly, because they are close to the majority of people in the network.
- Transport Networks:
  - Used to identify the most proximate stations or nodes that allow easy access to the rest of the network.
- Internet Networks:
  - Can be used to identify nodes or devices that may be central or vital for accessing data or other users.

Example of Closeness Centrality

Let's assume a simple network

```
A -- B -- C
|    |    |
D -- E -- F
```

If we want to calculate the closeness centrality of node "B":

The distances from "B" to other nodes are:
B to A: Distance 1
B to C: Distance 1
B to D: Distance 2
B to E: Distance 1
B to F: Distance 2
Total sum of distances: 1 + 1 + 2 + 1 + 2 = 7
The closeness centrality of B is . C(B)= 1 /7
The smaller this value, the less central the node B is in the network in terms of its closeness.

Important points:

- Nodes with high closeness centrality can be considered "centers of influence" in the network, because they are able to quickly reach most other nodes.
- In very large or disconnected networks, closeness centrality may not always be useful, as the distances between some nodes may be infinite.

Betweenness Centrality

is a measure used in network analysis to identify nodes that lie in the middle of many paths between other nodes in the network. This type of centrality reflects a node's ability to control the flow of information or resources between distant nodes.[6]

Definition of Betweenness Centrality

The betweenness centrality of a given node is measured based on the number of times the node lies on the shortest paths between two pairs of other nodes in the network. The more paths a node lies on between other nodes, the higher its centrality.

Mathematical Formula

To calculate the betweenness centrality of a node, the following formula is used:

$$C_B(V) = \sum_{s \neq v \neq t} \frac{6(s,t \setminus v)}{6(s,t)} \qquad (2)$$

Where:

- CB(V) is the betweenness centrality of node V.
- ϭ(s,t) is the number of shortest paths between nodes t and s.
- ϭ(s,t\v) is the number of shortest paths between t and s that pass through node v.

Concept

- Brokerage: If a node lies on many paths between other nodes, it is a "broker" in the network. Nodes with high brokerage centrality influence the flow of information or connections between nodes.
- Shortest paths: These are paths with the fewest edges between any pair of nodes.

Importance of Brokerage Centrality

- Social Networks:
  - In social networks, nodes with high brokerage centrality are well-positioned to influence the flow of information between individuals. For example, an individual who resides at the center of a social network and has many connections with other individuals in different locations may be a "broker" in the network.
- Transportation Networks:
  - In transportation networks, nodes with high brokerage centrality can serve as vital hubs for communication between different locations. Disabling these nodes may reduce communication between other areas.
- Internetworks:
  - In the internet, nodes with high brokerage centrality can serve as key gateways between different sub networks.

A Practical Example of Betweenness Centrality:
Suppose we have a small network containing nodes A, B, C, D, and E.

```
A -- B -- C
|    |    |
     D -- E
```

To calculate betweenness centrality for node B:

The shortest paths between pairs in the network are:
From A to C: The path is A → B → C.
From D to C: The path is D → B → C.
From A to E: The path is A → B → E.
From D to A: The path is D → B → A.
In this example, node B is located on many paths between other nodes. Therefore, betweenness centrality for B will be high.

```
A -- B -- C
|    |    |
     D -- E
```

- Advantages of betweenness centrality
  - It helps identify nodes that can significantly influence the transmission of information in a network.
  - It can be used in social analysis to understand the people who connect different groups in a social network.
  - In technological networks, it helps identify the devices or servers that play a vital role in the transmission of data across the network.

Intermediateness centrality expresses the extent of a node's influence in connecting different parts of a network. Nodes that are located on many of the shortest paths between other nodes have high intermediation centrality, making them important in facilitating or controlling the flow of information in the network.

- Community detection

Community detection is a fundamental concept in network analysis. It is used to identify groups or "clusters" within a network in which nodes are more closely connected to each other than to the rest of the network. These clusters are typically defined as groups of nodes that interact or are more densely connected to each other than to other nodes outside the group.[7]

- ▪ Why is community detection important?

Cluster detection is useful for understanding the structure of a network and discovering patterns that may not be apparent when looking at the network as a whole. For example:
• In social networks, it can help identify groups or teams within the network, such as friends or individuals with similar interests.
• In internet networks, it helps segment the network into groups of devices or servers that engage in intensive communication.
• In transport networks, it helps identify hubs or key points that connect different parts of the system.
The algorithm employs in two iterative stages:

Locally improving modularity by moving nodes to clusters that increase modularity. Merging clusters into a new network and repeating the process. It is simple, efficient, and able of analyzing large networks, but it may have problems detecting small clusters.

3.2 Practical part

In practical part, MATLAB was used to measure the three factors, such as calculating centrality (degree, closeness, mediation)[11] a s shown in figure 1 :

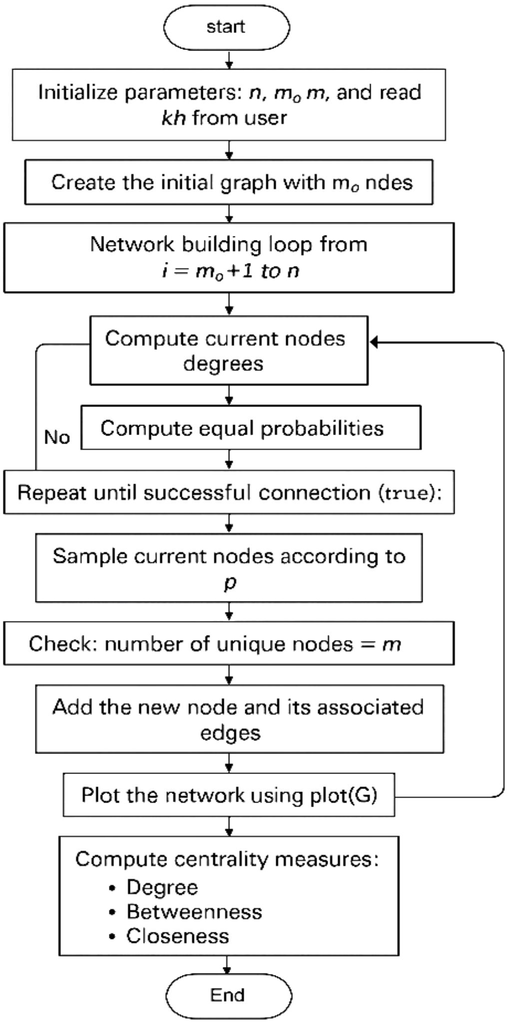

Figure 1. flowchart of measure the three factors

Such as using Louvain Algorithm , it employs in two iterative stages:

- Locally improving modularity by moving nodes to clusters that increase modularity.
- Merging clusters into a new network and repeating the process.

It is simple, efficient, and able of analyzing large networks, but it may have problems detecting small clusters.

```
A -- B -- C
     |
     D
```

4. RESULTS AND DISCUSSION

The analysis of the experimental network revealed that certain nodes play a essential role in the network structure. For example, nodes with high centrality degrees were the most impact in data flow, while betweenness centrality helped identify nodes that represent critical crossing points between network communities.

The Louvain algorithm was employed to detect clusters, demonstrating high effectiveness in uncovering the structure of community within the network.

The impact Factor is similar if the number of cluster is 3 ,it appeared that all clusters were equals.

TABLE 1. no. of nodes each factors

Degree Centrality	3	6	7
Betweenness Centrality	6	3	7
Closeness Centrality	6	3	7

Due to the small number of nodes, nodes 3, 6, and 7 exhibited similar importance, though some variations were observed.

it shown it's shown in next figure 2:

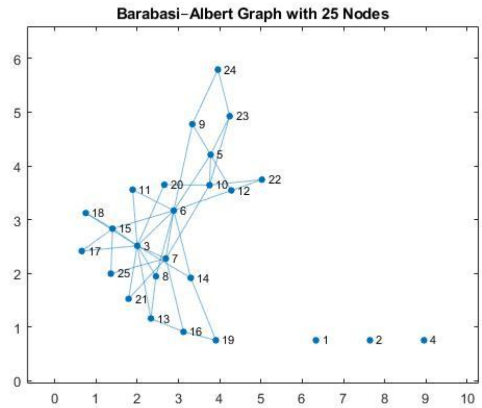

Figure 2. Barabasi-Albert graph with 25 nodes

In the second experiment   took 15 nodes and distributed them into Five clusters and the result is as Fellows:

TABLE 2. no. of nodes each factors

Degree Centrality	8	7	1	6	14

Betweenness Centrality	8	7	6	1	17
Closeness Centrality	8	6	1	7	13

It is evident that the ranking of influential nodes varies depending on the centrality metric used. example 14 is existant in Degree Centrality and not in the others.

and the same applies to 17 and 13 for Betweenness Centrality and Closeness Centrality respectively .

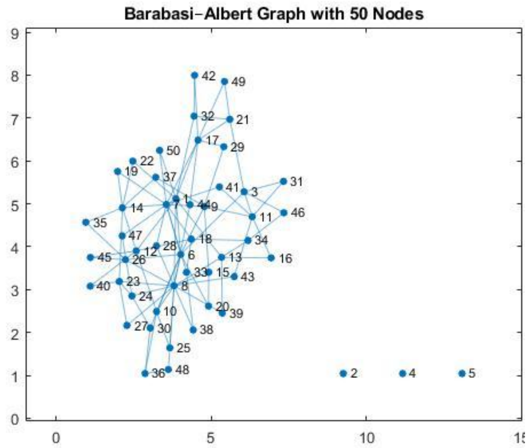

Figure 3. Barabasi-Albert graph with 50 nodes

In the third experiment seven nodes were used and distributed as follows:

TABLE 3. no. of nodes each factors

Degree Centrality	9	11	6	4	23	7	8
Betweenness Centrality	9	11	6	4	7	23	8
Closeness Centrality	9	11	6	4	7	8	5

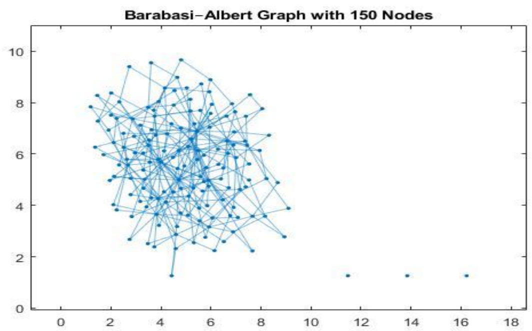

Figure 4. Barabasi-Albert graph with 50 nodes

There are different points.

Example 7 is in Degree Centrality and it is not in Betweenness Centrality and Closeness Centrality Also the Node 23 and 5 for Betweenness Centrality and Closeness Centrality respectively.

We can notice here that if network is bigger, then the difference will increase because the factors differ according to the mathematical rules of each Factor which determines the importance points.

the sort is important and it is shown in table 4:

TABLE 4. no. of nodes each experience

Experience	No. of nodes	No. of points	Factors		
			Degree Centrality	Betweenness Centrality	Closeness Centrality
1	25	3	3,6,7	6,3,7	6,3,7
2	50	5	8,7,1,6, 14	8,7, 6, 1, 17	8,6, 1, 7, 13
3	150	7	9,11, 6, 4,23,7, 8	9,11, 6,4,7 ,23,8	9,11,6,4,7,8 ,5

5. CONCLUSION

This study highlights the importance of using mathematical and statistical ways in analyzing big networks, focusing on MATLAB tools which provide a productive environment for pursuance algorithms and information analysis. Core nodes within the network were identified through different centrality metrics, and cluster discovery algorithms were successfully implemented.

Future work

The results of this study can be used in many practical fields. We can use this algorithm to measure large networks, and employ it in social network analysis, to identify the most influential individuals or entities on social media networks to target marketing campaigns, for example.

Practical application using MATLAB

This chapter explains how to apply theoretical concepts to analyze large networks utilizing the MATLAB environment. It includes representing a random network, calculating various centrality matters, and exploring clusters within the network by pursuance an algorithms utilizing MATLAB tools.

Representation of a Random Network

A random network of many nodes is made using the Barabási–Albert model or a random connection matrix.

Example: To create a network of 100 nodes with a 5% connection probability ,using MATLAB code:

```
n = 100;
p = 0.05;
A = rand(n) < p;
A = triu(A, 1);
A = A + A';
G = graph(A);
plot(G);
title('Random network representation');
```

Calculating Centrality Measures

using MATLAB functions to calculate the centrality of nodes:

-Degree Centrality
-Closeness Centrality
-Betweenness Centrality
-PageRank and Eigenvector

```
deg = centrality(G, 'degree');
close = centrality(G, 'closeness');
btwn = centrality(G, 'betweenness');
ev = centrality(G, 'eigenvector');
pr = centrality(G, 'pagerank');
```

Ploting the network with highlighted influential nodes

The significance of this study lies in its practical applicability across various domains.
The techniques covered in this paper can be extended to real data networks in diverse fields such as cybersecurity and public health.
By using the computational methods of MATLAB, the study provides an efficient framework for researchers to explore and analyze large data.

REFERENCES

1. Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review E*, 71(5), 056103.
2. Latora, V., Nicosia, V., & Russo, G. (2017). *Complex Networks: Principles, Methods and Applications*. Cambridge University Press.
3. Klein, D. J. (2010). Centrality measure in graphs. *Journal of Mathematical Chemistry*, 47(4), 1209–1223.
4. Gómez, S. (2019). Centrality in Networks: Finding the Most Important Nodes. In *Business and Consumer Analytics: New Ideas*.
5. Mbiya, S. M., Hancke, G. P., & Silva, B. (2020). An Efficient Routing Algorithm for Wireless Sensor Networks based on Centrality Measures. *Acta Polytechnica Hungarica*, 17(1), 83–99.
6. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
7. Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
8. Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
9. Flores, C. O., Poisot, T., Valverde, S., & Weitz, J. S. (2014). BiMAT: a MATLAB (R) package to facilitate the analysis and visualization of bipartite networks. arXiv preprint arXiv:1406.6732.
10. Santana, R., Armañanzas, R., Bielza, C., & Larrañaga, P. (2013). Network measures for information extraction in evolutionary algorithms. International Journal of Computational Intelligence Systems, 6(6), 1163-1188.
11. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

Arabic Abstract

تقدم هذه الورقة دراسة تحليلية للشبكات باستخدام تقنيات رياضية وإحصائية، مع التركيز على التطبيقات العملية باستخدام MATLAB. تهدف الدراسة إلى فهم البنية الداخلية للشبكات وتحليل العقد ذات التأثير الأعلى باستخدام عوامل المركزية، مثل مركزية الدرجة، ومركزية الوسيط، ومركزية القرب، بالإضافة إلى الخبرة المكتسبة باستخدام تطبيقات MATLAB لتمثيل الشبكات واستخراج المؤشرات الإحصائية.