A detailed study of different metrics based on closeness centrality

CSCI-6444 Introduction to Big Data Analytics

Spring 2022

The George Washington University

Isha Negi

G45705156

Abstract:

The fundamental precept of closeness centrality is the spread of information modeled employing shortest routes between the initial and final nodes. However, this approach might not be realistic in the diverse, dynamic networks prevalent in the graph modeling and information transmission scenarios in recent times. Although all researchers agree that closeness centrality is a significant structural characteristic of networks and that it is associated to important node properties and processes, there is no consensus on which nodal characteristic should be evaluated based on conceptual foundations, and there is very little agreement on the proper procedures for its measurement. This need for paradigm shift from restricted notions of closeness metric definition to encompass different dimensions or characteristics of nodal entities should be acknowledged in the contemporary network literature. The aim of this paper is to axiomatize the standard broad spectrum of closeness centrality measures into measures of related yet distinct characteristics. Seven metrics are discussed in detail - namely Information Centrality, Eccentricity, Radiality, Integration, Random Walk Closeness, Effective Distance Based closeness, Hierarchical closeness, which contribute to the existing work on axiomatizations of Closeness and establishing these as part of the same family measure. The key distinguishing feature for each of these associated family methodologies from specific closeness centrality metric, is also discussed in detail, which necessitates the development of a new associated closeness related measure altogether.

1. Background:

Several real-world situations incorporate data which translates to various types of interactions, processes or relationships among individuals, organizations, or groups, thus establishing large scale networks. Graph representation is used to transform these complexes inter relations to data form. Graph networks allow us to navigate through data in a flexible way and process the rich set of connections without the use of any pre-defined model. In Analytics, the graph data is modeled by using machine learning algorithms to analyze the data using statistical/data mining techniques to obtain hidden patterns or user behavior from the data. Transforming data to graph representation (mathematical structures) to obtain network data to find important node behavior or node interactions using various analytic techniques and algorithms is a challenging task. The emergence of graph theory can be traced back to as early as 1735 when the Swiss mathematician Leonhard_Euler solved the Königsberg bridge problem.

A graph is a mathematical abstraction of relationships that emerge in nature; hence, it cannot be coupled to a certain representation. Many graph measures and metrics include vertexes/nodes which denotes the important entities in the graph data and edges that indicates the relationship between two nodes. Centrality is one of the most studied concepts in network analysis.

With the advent of dynamic complex graphs in big data, it is imperative to recognize important entities within the network which can further help in exploring the characteristics of the network. The process to do a rigorous study of each node in extensive data is cumbersome, time-consuming as well as costly in real-time. Hence, it necessitates the use of a quantifiable metric that can gauge the influence or effect of the node with respect to others.

Closeness centrality concept was limited to direct ties or associations of the node with its immediate or nearby neighbors which would be inefficient in case of complex, diverse or dynamic graphs. Graph databases are responsible to address challenges in wide range of domains nowadays which involves heterogeneous, complex data with multiple linked relationships. It is imperative to examine the indirect contacts too which are instrumental in channeling information or influences within the network. This purpose can be achieved by computing various metrics which provide a means of assessing behavioral correlation with a network measure based on chain of contacts or nodes.

2. Centrality:

Centrality is one of the most important fundamental concepts that is used to identify the critical entities/nodes from the structural point of view. Centrality algorithms is one of the traditional categories of algorithms used in graph analytics. The term ‘centrality’ does not limit to the topological origin or reference. Centrality denotes to the prominent nodes present within the network where the importance of a particular node can be determined depending on one or combination of factors. The ‘importance’ of a node varies which can be established depending on the definition perspective. Few nodes are vital in comparison to the others in the network structure as they tend to possess higher status in terms of function or structural orientation. A node can be deemed important if it has lot of direct connections, holds a key infrastructure position where it is transitively connected to other important nodes or is available on the shortest path of lots of node pairs. The importance of assessment of network centralization can be verified by the experimental results of removal of central nodes in social network [10]. The results signified the whole network centralization is decreased sharply with removal of important central nodes. Centrality- based ranking of nodes in real-world networks is purely a link- statistic based approach and does not depend on any offline information [1]. The node degree and the shortest path between nodes are the most informative values available to determine advantage of a node with respect to its neighbors. Degree, betweenness, closeness (as local measures) and eigenvector (as global or spectral measure) are the most used centrality measures [2]. These measures can be calculated using the knowledge gleaned from network properties and network topology. Each of the mentioned measure above has its own context based on functionality aspect during analysis or the anatomy of the heterogeneous network being analyzed.

Traditionally, networks were represented with the use of adjacency matrix. Adjacency matrices are ineffective for large scale networks as it is sparse and high dimensional. Network embedding, which aims to learn low-dimensional continuous representations for nodes, encompasses techniques for mapping graph nodes to vectors of real numbers in a multidimensional space. It assumes that the similarity between nodes in the network should be reflected in the learned feature representations. The vectors can then be used as input to various network and graph analysis tasks, such as link prediction or node classification. A good embedding is expected to preserve the structure of the graph, but it somehow ignores the important concept of centrality.

A model was proposed in [16] to use the traditional network embedding part to preserve the network structure as well as incorporate centrality information to comprehend node representations. The focus is on preserving the ranking of the centrality scores instead of the absolute values in order to retain the centrality information of the network. The listwise ranking algorithm listMLE was adopted to maintain the ranking among the nodes. The below problem statement was evaluated to address the node representation in embedded networks:

Given a network G = {V, E} and a centrality measure function f, we want to learn a set of representations U for the nodes and a scoring function f' to map the representations to score which can preserve the network structure as well as minimize the difference between $\{r_1,...,r_N\}$ and $\{r_1',...,r_N'\}$ [16].

Here, V is a set of nodes and E is a set of edges between those nodes. The nodes $\{v_1, . . ., v_N\}$ can be sorted according to the centrality scores $\{f(v_1),...,f(v_N)\}$, let the ranking list be $\{r_1,...,r_N\}$, where r_i denotes the rank of the node v_i. The set of nodes can be also represented as $\{v(r_1),...,v(r_N)\}$ using the ranking as index. Let $U = \{u_1,...,u_N\}$ be the corresponding representations for the nodes, f' be the function that maps the representations to the centrality scores based on the representations, and $\{r_1',...,r_N'\}$ be the ranking list based on f'. The representations U and the function f' need to be learned [16].

An application of the centrality measures was practically implemented to do a comprehensive study as well as centrality analysis of the contiguous states of the Unites States [3]. It fundamentally used the concept of the four measures related to centrality: degree, closeness, betweenness and eigenvector to rank the states attributed to the location of respective state. The research of the connected graph of the states within the country was thought to serve the purpose of using the extracted information to design road/rail transportation networks in the future based on landscape and geographically centralized locations.

2.1 Degree Centrality

Degree Centrality denotes the count of the number of links or edges from a vertex i.e., v(m) over its maximum value [2], i.e. (N-1):

$$C_D(m) = \frac{\sum_{i=1}^{N} a(v(i), v(m))}{N-1} = \frac{\deg(m)}{N-1}$$

[2]

where N is the number of vertices or network size and a(v(i), v(m)) is the corresponding element of adjacency matrix of the network graph G = (V, E). It is the most basic and intuitive measure of centrality. In this measure, the calculation involves the edges that directly connects to the adjacent vertices/nodes. It is a local centrality measure as it only accounts for the immediate surrounding nodes of the individual node, irrespective of the size of the network. Since degree and betweenness centrality indices are used to measure the center position and function of nodes in networks, the correlation between the two was estimated for practical significance in social networks using Pearson correlation coefficient [10]. The results demonstrated positive relation between the two indices with a higher degree node showing high betweenness centrality.

2.2 Betweenness Centrality

Betweenness Centrality denotes the number assigned to each vertex that measures the number of times it falls on the geodesic path. Geodesic path refers to the shortest distance in terms of edges or links between two vertices in a network.

$$C_B(m) = \sum_{s \neq m \neq t \in V} \frac{\sigma_{st}(m)}{\sigma_{st}}$$

[2]

where σ_{st} is the number of shortest paths from node s to node t, and σ_{st} (m) is the number of shortest paths from node s to node t that pass through a vertex m [2]. Any vertex with high betweenness centrality measure indicates higher value or importance depending on the control of information flow it has in the network considering that the flow of data takes the shortest route. Unlike degree centrality, betweenness is defined as a global centrality measure as it provides a value taking the whole network into account and then providing a normalized or scale value. Since it involves each node to compute the shortest path between all pairs of vertices on a graph, it is not considered as a time or memory effective process.

2.3 Eigenvector Centrality

Eigen centrality is a measure of the influence of a node in network. It is similar to degree centrality as it also depends on the neighborhood of each node. However, it does not only take the number of immediate neighbors into consideration but also how important those neighbors are. The approach follows the assignment of relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes [5]. A high eigenvector score denotes that a node is connected to many nodes who themselves have high scores. The relativity factor is administered where a particular node connected to an important node will be deemed as a more significant node in the graph. The eigenvector centrality calculation is written with the following equation:

$$C_E(v_i) = \frac{1}{\lambda} \sum_j A_{ij} C_E(v_j)$$
$$Ax = \lambda x$$

[5]

where C_E is the eigenvalue centrality value calculated for vertex v_i with eigenvalue λ using an adjacency matrix A and matrix of eigenvector centrality.

2.4 Closeness Centrality

Closeness centrality is defined as the overall mean of the shortest path between a vertex m and all other vertices reachable from it [2]. It is a measure of the number of hops on the

shortest paths from the vertex to every other vertex on the graph.

$$C_c(m) = \frac{\sum_{t \in C \backslash m} d_G(m,t)}{M-1}$$ [2]

where M is the size of the network's connectivity component C reachable from vertex m

It assesses the communication influence of a specific vertex within the network in comparison to the average path length of the whole network. Both these measurements use the computation of all-pair shortest path distances in the network based on the breadth-first search (BFS) method [6] at every vertex. In addition to the locality of reference for a particular vertex, the structural formation of the graph network as a whole also needs to be considered in case of closeness measure. Proximity or closeness of a node to all the other points present in the network is responsible to determine the independence of that point. A point with larger closeness means that it can more easily avoid the control potential of other points, in other words, the point is more efficient.

Closeness centrality measure for a particular node can be assessed to evaluate different aspects in real-time situations – whether a node is an important boundary spanner in knowledge graphs, key or distinct node in a network, failure prone nodes in a contribution network, authority experts or information propagators in social network analysis, or node actors responsible for low data transmission delay in sensor networks [6]. On the contrary, the average path length is used in situations when the whole topology of real world needs to be analyzed like World Wide Web, estimate the transfer efficiency between all metabolite pairs in a metabolic network or to measure time needed for evaluating the function of binary decision diagrams. A major benefit of closeness centrality is that it indicates nodes as more central if they are closer to most of the nodes in the graph. This strongly corresponds to visual centrality—a node that would appear toward the center of a graph when we draw it usually has a high closeness.

3. Metrics derived from the concept of Closeness Centrality

3.1 Information Centrality

Social and communication graphs abound in different domains in daily life to share and receive ideas, opinions and beliefs. Communication in these scenarios can range from engaging in gossips, email delivery, sending text messages, updating social media statuses to navigating the world wide web. Different types of data flow process take place in social networks and hence, information flow or propagation is subject to inherent influence. Networks can differ as far as how node-to-node transmission occurs or how a path through the network is taken, producing various types of information flow processes. When metrics are applied to an inefficient flow process, the outcome can be misleading and frequently wrong. The initial network literature work acknowledged four general measures of centrality termed ‘degree’, ‘closeness’, ‘betweenness’ and ‘eigenvector’ which was motivated by the structural properties of the nodes in a graph. The first three mentioned key centrality concepts were partially motivated by the structural properties of the center of a graph which is dependent on the basic idea of point centrality. Point Centrality of a node in a graph is the adjacency count of its constituent nodes. Degree Centrality is only concerned with the nodes in direct contact. Closeness and Betweenness centrality are measured based on the geodesic or shortest path concept which is dependent on local pair dependency. Since both theses metrics consider geodesic pathway as an integral deciding factor of selecting central nodes within a network, the same structural points of the network are included.

Central nodes are deemed to be vital as they have the capacity and reach to influence maximum elements within the network. Finding the centralities of nodes based on their information gain, which takes into consideration the information gains of their neighboring nodes as well, is one of the most effective techniques to locate prominent nodes in a network [20]. However, this theoretical choice often neglects other reachable, non-geodesic pathway nodes responsible in the flow of communication. Geodesic paths not necessarily capture the same information as in the circuitous route of random communication or through peripheral intermediary nodes. Eigen vector measure associated with the adjacency matrix was another critical centrality metric devised in further research. Although it is considered an important characteristic of network, it fails to consider multiple paths shared between two nodes. The basic centrality measures fail to understand very large networks due to their respective limitations to include all the possible paths. This, in turn, requires for decomposition of the large networks into smaller units or components. However, this approach also compromises the subtle network infrastructure for computational purpose. The appropriate usage of centrality metrics depends on an understanding of the type of network flow [15].

It is essential to recognize the central actors or nodes in information flow networks to develop an understanding the coordination and communication between them as well as who is governing the flow. A contemporary measure to assess all the paths in the non-directional network was introduced which assigned a relative weightage or ranking of the nodes as a function of the information the nodes contained. An information centrality value of a node is defined as the average of the information of all paths originating from that node [15]. Since information is the reciprocal of the path length (reciprocal of variance, according to the theory of statistical estimation), this metric is related to closeness. It can be, therefore, deduced that the information in a path is the inverse of the length of a path. When there are two or more routes linking nodes that contain some of the same incident nodes, the information is computed by inversion of a symmetric matrix that contains the number of incident nodes or points the routes contain. This measure was technically defined but not calculation intensive which made it possible to be used in investigation of networks in practical applications. The information centrality for a node is given by:

$$\overline{I}_i = \frac{n}{\sum_{j=1}^{n} \frac{1}{I_{ij}}}$$

[15]

where n is the number of nodes and I_{ij} is the centrality of a path from node i to j.
Borgatti in his research arrived at a typology containing two dimensions along which different flow processes in a social network can vary [15]. One of the dimensions involved

the node-to-node transmission in the form of broadcasting, parallel or serial duplication and transfer. Unlike transfer, the other transmission modes enable copying as well as storing the information before passing it on to the next node. The other dimension considers the course taken in the case of the network where the information relaying can be the shortest route, called a geodesic, or it can be a path, trail or walk. Paths do not repeat nodes and links, trails do not repeat links and in walks both nodes and links can be repeated [15].

In [15], a simulation of the network and the information traversal within it was used to determine the applicability of information centrality in social networks. At each node, a variable named ‘information’ was utilized that was only incremented for replication and not for a transfer. Certain statistics were gathered by counting the number of times a piece of information goes through a node or the time it takes to reach a node. The frequency of arrival refers to the number of times information flows through the intermediary node that is not the destination node. The arrival time is the amount of time it takes for a node to be reached by all other nodes. A greater frequency of arrival value implies a more central node, because the node is frequently reached. A lower value for arrival time signifies a more central node, since a node is much more central if it takes less time to get to it. This led to the conclusion that a node that takes a longer time to reach is often not a very prominent node. It is based on the assumption that information centrality, like closeness centrality, is concerned with the amount of time it takes traffic to reach the node (the information it receives), rather than the actual frequency of traffic passing through the node. This is also why information centrality is referred regarded be a measure of proximity. Another viewpoint is that information is delivered to a node once it is reached, and subsequent transfers do not contribute to the information. The ranking of nodes based on simulated findings were compared to the ranking of the estimated information centrality of the network's various nodes. The result indicated that the information centrality measure correlates with central nodes when the information is conveyed through a walk where nodes or links can be repeated. The outcome demonstrated that Information Centrality is more like Eigenvector and Degree centrality than to Closeness centrality as postulated by previous research [15. Information centrality does not correspond with "arrival time," as we previously assumed, but rather with "frequency of arrival." Closeness Centrality, on the other hand, as predicted, corresponds more with GeoArT (the transfer through Geodesic traversal).

A key challenge is to find a route that can provide acceptable delivery performance and minimal end-to-end delay in a disconnected network graph where nodes may move freely. This challenge was addressed in [17] with the research proposal of a social network analysis metric that might be utilized to enable an innovative and realistic message forwarding mechanism in disconnected delay tolerant MANETs. The information metric was used in [18] to propose a generic information spreading forensic model and establish a novel methodology for estimating heterogeneous spreading rates, source start time, and information source location using sequential and dependent snapshots. Another implementation of information measure of complex network theory was used in [19] which offers a comprehensive approach of emergency evacuation traffic organization which integrates an emergency evacuation corridors optimization method with traffic organization method of emergency evacuation corridors.

3.2 Eccentricity

Most of the current evaluations for node importance in complex networks can not reflect the reality of some real-world problems objectively [7]. The statuses of nodes in real world data are not consistent and therefore, network analysis is often accompanied by structural constraints. It necessitates the proposal of utilizing eccentricity perspective of node to recognize as well as find the center and median of graph. Eccentricity, being a distance-based and global reference measure, is like closeness centrality. However, it can be considered a much simpler notion than closeness while administering the optimal nodes in terms of location by computing vertex node centrality.

The eccentricity $\varepsilon(i)$ of a vertex i is defined as the length of the longest shortest path from i to any other vertex in the graph. The exact eccentricity of all vertices in an unweighted graph can be ascertained by conducting a breadth-first search (BFS) from all vertices in a graph. However, for large-scale networks, the computational time of such a naive technique is untenable. The eccentricity of u, defined as the longest shortest distance between from one node u to all other nodes in the graph, is composed of two core attributes of the entire graph

— diameter and radius. Over all the nodes in a graph, diameter is the maximum eccentricity and radius is the minimum eccentricity.

Given a node u of a graph G (V, E), the eccentricity of u is defined as:

$$ecc(u) = \max_{v \in V} dist(u, v). \quad [8]$$

The below figure illustrated in [8] describes the nodes on the graph G, labelled with the computed eccentricity of each node in respective squares**.** The color gray scale indicates, for each node, the eccentricity value: a node with a darker color means that it has a smaller eccentricity. For example, the eccentricity of node v1 is calculated as $ecc(v_1) = \max_{v \in V} dist(v_1,v) = 2$, and $ecc(v_{11}) = \max_{v \in V} dist(v_{11}, v) = 4$. Intuitively, a node at the center has a smaller eccentricity than the node at the border.

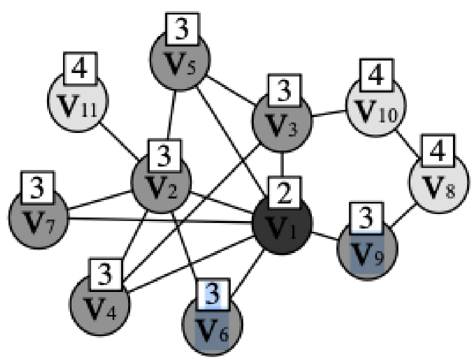


Fig 1: Node Eccentricity Graph (Qin Qion,2016)

The eccentricity measure can be regarded as the ease with which all other nodes in the network can access a node's functionality. As a result, a node with a high eccentricity compared to the network's average eccentricity will be more easily influenced by other nodes' activity, or conversely, could easily influence several other nodes. A low eccentricity compared to the network's average eccentricity, on the other hand, could appear to suggest a marginal functional role (although this should be also evaluated with other parameters and contextualized to the network annotations).

Vertex eccentricity is a well-studied and relevant measure of vertex centrality, with applications in tissue characterization and categorization in biology as well as assessing chemical compounds [9]. By implementing a breadth-first search (BFS) from all vertices in a graph, the absolute eccentricity of all vertices in an unweighted graph can be estimated.

3.3 Radiality

Network analysis primarily focuses on role structure of the graph to investigate the role relations within the networks. The purpose of centrality metrics is to investigate the positions of the nodes and its respective influence on the overall network. The centrality of a node is a quantification of its importance in a graph by considering its various structural properties as defined by the relationships or edges. In an effort to contribute to the existing centrality literature, new centrality metrics were introduced in the field of network analysis. One such metric was proposed in [11] to gauge an individual node's radiality to measure how well its ties reach out in the network. Radiality refers to the degree an individual's relations reach out into the network providing access to many and diverse others [11]. It was termed as a substantive value as measure of reachability which could be measured easily with standard graph-theoretic tools. Radiality can be viewed an extension to the concept of closeness centrality which aims to provide an enhanced explanation of the associations in the network in terms of nodal characteristics.

Radiality is an interesting concept which was proposed to visualize node centrality information in the context of the overall graph structure which is captured through intervertex (graph theoretical) distances. Intervertex distances can be addressed as the computation of the shortest path lengths, also known as geodesic paths, between two nodes in unweighted and undirected simple graphs. Traditionally, the prior centrality metrics used geodesic and its reciprocal to compute centrality. On the other hand, radiality uses the reversal of the geodesic distance measure which is conceptually similar to the reciprocal but computationally different. The reversal of geodesic provides a linear transformation of data that is consistent with the original metric. Radiality is the degree of an individual which describes its ability to form ties in the network to form novel paths for traversal of information within the network. A node which has direct contact with other nodes not linked directly with each other has a more radial network than the node having contacts with others which are connected to each other. High individual radiality means that it takes fewer steps, on average, for that node to reach everyone else in the network through its outward-bound nominations sent [11]. Higher and lower values in terms of radiality can be deemed to be more meaningful when compared to the average radiality of a graph which can be computed by averaging the radiality values of all the nodes in the graph. High radial value corresponds to higher degree of node proximity whereas lower radial value determines the peripheral nature of the node. Unlike closeness centrality, radiality is not definitively informative on the centrality of the node and is only considered to provide an average tendency to node proximity or isolation. A node having high eccentricity along with high closeness value and high radiality tends to hold a high central position in the graph.

A classic application of radiality is in biological networks, specifically protein interaction networks, where the measure of possibility of a protein to be functionally relevant for the other proteins in the network had to be determined [13]. The identification of influential or essential proteins which play critical role in cell processes is feasible for biomarker discovery, drug repurposing or drug design. An examination of centrality profile of nodes was done in [13] to predict and prioritize the dominant influential proteins where radiality came up as one of the most significant indicators in the results. It can be inferred that a biological signaling network with very high average radiality can be is more likely organizing functional units or modules, whereas a network with low average radiality will behave more likely as an open cluster of proteins connecting different regulatory modules.

3.3.1 Radial Layouts

Visual rendering of graphs offers an effective medium of encoding the underlying data while maintaining the structural characteristics of the entities in the network. In conventional radial layouts, the distance of nodes from the geometric center (or origin) of the layout depends only on the node’s centrality. It is based on the fact that the nodes having a higher centrality value will be placed closer to the central node being considered as origin in the layout while forming rings or concentric circles in visual representation. A method in [12] was proposed with the help of visualization using layouts of the position of nodes in the graphs. The primary aim of layouts was to gain insights from the graph structured data for further research related to the network data. Decoding the network data in terms of layouts serves multiple purpose ranging from reducing the clutter and edge crossings to conveying the hierarchy within the nodes in terms of importance or usefulness.

Centrality indices are typically described as the real-valued functions over the nodes of a graph. While the emphasis of the various centrality metric definitions can be different, they all share a common characteristic of depending only on the structure of the graph rather than parameters associated with the nodes [14]. Data depth is an integral notion which is relevant to the underlying technicalities of a network as it characterizes the generalization of data depth to the vertices on graphs. Data depth is a family of methods from descriptive statistics that attempts to quantify the idea of centrality for ensemble data without any assumption of the underlying distribution [12]. Graph centrality is a type of data depth on the nodes of a graph which can illustrate the properties of the node along with layout methods. Data depth methods are responsible for exhibiting properties like maximum value at geometric center, zero at infinity and radial monocity.

3.3.2 Enhancement of Radial layout- Anisotropic layout

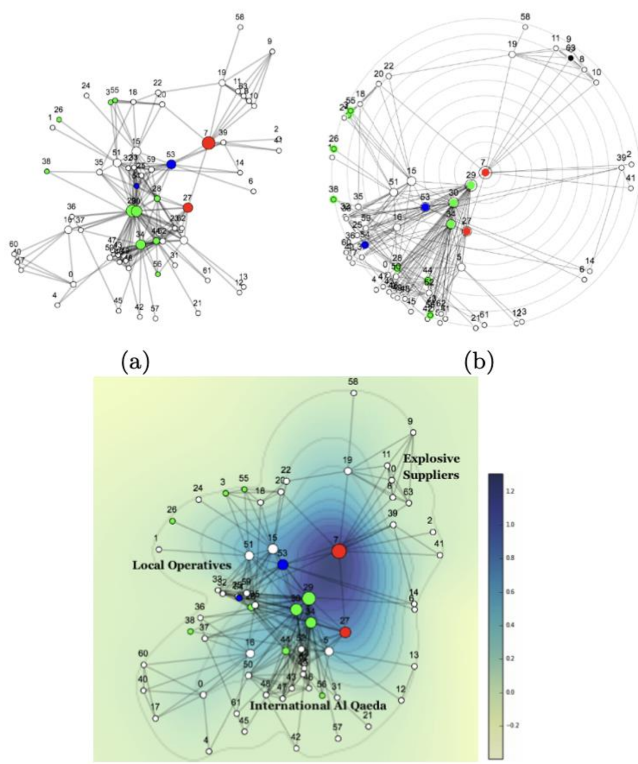
Various methods or approaches have been explored to determine the radial layout for a graph with associated entities. While minimizing the edge crossings by deletion is considered apt for discrete centrality values, optimizing a stress energy by including a penalty for representation error (of graph distances) as well as deviation from radial constraints is followed to tackle continuous centrality values. Preservation of distances and Anisotropic radial monocity are devised as the two important criteria while determining the layout of a graph. The Preservation criteria administered that the measurable geometrical distances should be approximate to the theoretical distances between the nodes in the graph. This was done to ensure that the results of the computation were comparable to the visual and theoretical graph structure. On the other hand, radial monocity ensured that the symmetry of the layout is maintained and translated effectively in the visual format. It should convey that along the ray travelling away from the central node position, the nodes with a lower centrality should be placed geometrically further along the ray. A new layout algorithm was discussed in [12] to administer that closed curves are more general constraint than the previously proposed circles which can afford to provide more flexibility to preserve vertex relationships compared to existing radial layout methods. General layout methods like multidimensional scaling fail to readily convey whereas the associated circular centrality constraints in the state-of-the-art methods for radial graph layout make it difficult to preserve the holistic structure of the graph. The new proposed layout method attempts to relax the constraint that requires nodes with similar centrality to lie on a circle, and instead, allow for such nodes to be constrained by a more general shape: a simple closed curve or centrality contour. Centrality contours are nested isolevel curves on a smooth, radially decreasing estimate of node centrality values over a 2D field [12].

Multidimensional scaling (MDS) is attributed as an objective function that aids in visualizing the similarity (or dissimilarity) between members in a data set. MDS has been the foundation for a range of graph drawing algorithms that aim to achieve an isometry between graph theoretical- and Euclidian distances between nodes. The proposed approach in [12] modifies the multidimensional scaling (MDS) stress to include the estimation of a vertex depth or centrality field as well as a term that penalizes discord between structural centrality of vertices and their alignment with this carefully estimated field. A modified metric MDS with distance scaling has been successful in achieving additional flexibility in placing the nodes afforded by the centrality contours over circles, in conjunction with some additional visual cues in the background associated with a better trade off than the existing methods in conveying centrality and general structure together. In the context of graph drawing, given a distance matrix based on graph- theoretical distance, the goal is to find node positions $X = \{x_I : 1 \leq i \leq n\}$ that minimize the following sum of squared residuals—also known as stress:

$$\sigma(X) = \sum_{u,v} w_{uv} \left(d_{uv} - ||\bar{x}_u - \bar{x}_v||_2 \right)^2,$$

[12]

where $wuv \geq 0$ is the weighting term for residual associated with pair u, v. In the proposed work we employ a standard weighting scheme for graphs, known as elastic scaling, by setting $w = (d_{uv})^{-2}$. This gives preference to local distances by minimizing relative error rather than absolute error during the optimization. Here, the multidimensional scaling (MDS) stress is modified to include the estimation of a vertex depth or centrality field as well as a term that penalizes discord between structural centrality of vertices and their alignment with this carefully estimated field [12]. The key implication of the anisotropic centrality field in this method is that more central nodes are allowed to be placed further from origin than less central nodes—without an energy penalty—if they do not lie on a common ray, which aids our objective of achieving a better balance between visual representations of centrality and structure than possible with existing methods.



(a) (b) (c)

Fig 2: Network of terrorists and affiliates connected to the 2004 Madrid train bombing using (a) MDS, (b) radial layout, (c) anisotropic radial layout (Raj Mukund, 2017)

Few known social networks using the anisotropic radial layout visualization are discussed in detail in [12]. These include the Zachary's karate club for network of friendships, Terrorist network from 2004 Madrid train bombing and coappearance network for characters in Les Misérables novel. The successful attempt of ARL (Anisotropic Radial Layout) to preserve the structure and cohesiveness of the core influential actors or nodes in comparison to radial layout which obscures the internal structure of the network is effectively indicated in the visual drawings. ARL could illustrate the distinguishable locality of various clusters while highlighting the intermediaries of high centrality value accountable for acting as a bridge between various groups.

3.4 Integration

Most complex networks are heterogeneous and large-scale, with a wide range of entity types and relationships, and a community structure that is often characterized by overlap, complexity, and diversity. Centrality measures are scalar values assigned to each node in a network in order to assess its significance based on a hypothesis. The four primitive centrality measures were deemed the fundamental basis which was sufficient to study network theory. After thorough research, two new centrality measures to the four fundamental metrics of positional and structural equivalence as ways to tap into the role structure and influence of nodes in a network: radiality and integration.

Integration is a metric that indicates how well an individual is linked in a network. Individual 'integration' refers to the degree an individual is connected to many and diverse others in a network. Integrated individuals, all other things being equal, should have earlier access to information since they are, on average, closer to everyone [11]. Network interconnections and chains of associations are instrumental in dispensing direct access to resources and behavioral impacts within the network nodes. However, most structural assessments of informational and behavioral flows rely on positional and/or centrality measurements, which do not directly quantify these chains of linkages. Indirect relationships are critical not just for a central actor's network manipulation, but also because they are the portals through which ideas, influences, or information socially distant beyond his social circle might reach him. The fewer indirect links one has, the more isolated one will be in terms of knowledge of the world outside of his own social circle; consequently, it is critical to bridge weak associations. As a result, integration and radiality as indicators of connection and reachability respectively may be advantageous. Diversity metrics such as personal network density (Scott, 1991) and heterogeneity (Marsden, 1987) are widely used to identify an individual's immediate environment [11]. Although density and heterogeneity are crucial in understanding a person's network integration, they are confined to personal or centric research and do not go beyond first-order nominations.

Integration successfully enables to quantify diversity of ties extending beyond first-order relations. Integration metrics make use of geodesic path to compute the nodal characteristics; similar to radiality. The results have shown that integration is correlated with in-degree and radiality is correlated with out-degree. It is possible to scale the number of zones utilized in the integration measure by varying the distances used as cut-offs values of ‘meaningful' reachability, allowing the researcher flexibility over the amount of network influence he or she wishes to include in the measure.

The degree to which an individual's inward nominations incorporate him or her into the network is measured by integration. The initial point integration measure proposed was built on the works of Guimares' (1972) and is operationally defined as:

$$I(k) = \frac{\sum_{j \neq k} \mathrm{RD}_{jk}}{N - 1}$$

[11]

where I(k) is the integration score for node k, RD_{jk} is the reverse distance computed from the geodesic between actors j and k, and N is the network size [11]. The geodesic distance is found, and then the geodesic distance is subtracted from one plus the maximum value within the geodesic matrix to get RDjk (i.e., one plus the diameter of the network). In the reversed distance matrix, the distance values for nodes that are not accessible from one another, as well as the values on the diagonal, are set to zero. Higher l'(k) values suggests that the individuals are approachable in fewer steps on average than those with lower l'(k) values.

The metric is proportional to the network's size and diameter, therefore a reverse score of l'(k) indicates how distinctive one individual is from everyone else in the network in terms of the diameter, the longest geodesic.

The key distinction between integration and closeness centrality is that integration constructs a modified closeness measure by reversing the distances between nodes and then averaging these values for every individual. Despite the fact that both proximity and integration measurements are theoretically comparable, a slight computational difference between reversing and reciprocating results in distinct values for both metrics for a given node.

Integration may be performed on an asymmetric matrix, yielding a directed, distance-based measure of centrality which can be treated as an improvised closeness centrality as it was theoretically considered studied for undirected graphs.

Following the reasoning of Freeman (1979) and Wasserman and Faust (1994), an overall integration/radiality index for the network can be determined [11]. Network integration is computed with the following formula:

$$I = \frac{\sum_{k}[I^{*} - I'(k)]}{\max \sum_{k}[I^{*} - I'(k)]}$$

[11]

where I^{*} is the maximum relative integration value observed in the network and the I'(k) are the other observed relative integration values in the network [11]. The numerator is the difference between the greatest relative integration score and the rest of the network's values. One individual has maximum integration as an inward pointed star in a network of size N, with the center node having a relative integration score of 1, while the remaining nodes have a value of zero.

The research in [22] provides an integration centrality-based methodology to investigate the connections between stations in a streamflow monitoring network and determining the relevance of individual stations in the Pyeongchang River basin in South Korea. The study findings confirm that the stations along the main river channel in the basin's middle have a high degree of centrality, whereas tributary stations have a low degree of centrality. With community-based clusters, integrated centrality is utilized to assess of the value of streamflow stations to develop maintenance strategies that are both effective and efficient. The constitutive feature of structural integration within the network was illustrated with an example of viral marketing in [21] where the integration of well-connected individuals was described to be of considerable importance to draw the attention of the potential larger audience to a brand, a product, or a campaign .The integration of members who occupy a vital role within their network is also of substantial use in product creation and particularly in the detection of trends, because these actors have access to information about a range of other actors.

3.5 Random Walk Closeness

Sleep modes, channel variations, mobility, device failures, and other factors create dramatic structural changes in many wireless and mobile networks. Topology-driven algorithms are ineffective for such networks because they need a lot of overhead to keep topology and routing information up to date, as well as recovery procedures for crucial sites of failure. Classical network measures do not display characteristic features like the small world and clustering property which leads to the emergence of new metrics to comprehend the structure underlying the heterogeneous networks. Random walks are beneficial in such cases since they don't require explicit knowledge of network architecture. In the interest of robustness, random walks quantify how central a node is located regarding its potential to receive information randomly diffusing over the net- work.

While the shortest-path distance between two nodes can change dramatically with the insertion or deletion of a single edge, random walks distances account for multiple paths and provide a more global view of the connectivity between two nodes. With this spirit, one node's random-walk centrality in proportion to the rest of the network is defined as the estimated time required to come across this node in a random walk that starts anywhere in the graph. In highly dynamic networks, where effective topology computation is essential for enhanced network performance, random walk-based techniques are especially beneficial for finding packet delivery subgraphs. A random walker running on a network gives a sequence of vertices by selecting a random neighboring vertex from each vertex. While describing centralization of information wandering throughout networks, the RWC connects structural heterogeneity to asymmetry in dynamics.

The mean initial passage time is an essential feature of random walk (MFPT). It is determined by the number of links in the shortest path between two nodes, and both forward and backward motions are symmetric, unlike the asymmetric motion of a random walk. The difference in MFPTs characterizes this asymmetry, and the difference is further defined by a potential-like quantity known as random walk centrality (RWC). The estimated number of steps it takes for the process to reach node j from node i for the first time is the mean first passage time from node i to node j:

$$H(i,j) = \sum_{r=1}^{\infty} rP(i,j,r)$$

[26]

where P(i,j,r) signifies the probability that getting from I to j takes exactly r steps for the first time.

Consider a weighted network, either directed or undirected, with n nodes represented by j=1,..., n, and a random walk process with a transition matrix M on this network. The random walk closeness centrality of a node i is the inverse of the average mean first passage time to that node:

$$C_i^{RWC} = \frac{n}{\sum_{j=1}^{n} H(j,i)}$$

[26]

In [24] an enhanced version of random walker is discussed in which the next neighbor is chosen using a decision parameter based on the centrality of the node in the network. The path length is greatly reduced using these choice-based random walk algorithms. Load balancing in random walks on graphs refers to distributing the number of visits per node as evenly as feasible. The goal is to traverse the graph in such a fashion that all nodes receive nearly the same number of visits at each cover step. In energy-constrained wireless networks, where such protocols may be deployed, load balancing is of critical importance. The random walk with choice, RWC(d), is a well-known variation of the random walk that takes a subset of d neighboring nodes before deciding to move to the node that maximizes the value of a given metric; this metric captures the number of (past) visits to the node [25]. The ERWC(d) approach is used to define a metric for every node that captures not only the actual visits to the node, but also the intensity of the visits to the node's neighborhood. The simulation results in [25] demonstrate that the ERWC(d) outperforms the RWC(d) in terms of cover time, maximum node load, and load balancing at cover time, primarily on random geometric graphs and 2-dimensional torus.

The examination of an economy's input-output model, which is represented by a densely linked weighted network with significant self-loops, is one of the most important applications of random walk centrality in the area of economics [27]. It revealed key components of diverse national economies while identifying a central node as a sector that is most likely to be immediately or significantly influenced by a random supply shock.

3.6 Effective Distance Centrality

Several strategies have been investigated by studying the topic of influential or important node ranking from the perspective of performance evaluation. A novel effective distance-based centrality (EDBC) algorithm for the identification of influential nodes based on certain criterion in concerning, complex networks. EDBC comprises effective distance, K-shell, potential-like characteristics. EDBC can be termed as aa robust algorithm consisting of numerous characteristics attributes that have been inspected on unweighted networks and structure in order to sort out the important nodes. The postulated EDBC is parameter independent and unconventional, as it does not rely on historical knowledge or parameter alterations. EDBC, which is based on two-stage neighbor nodes strategy, decreases the computation overhead. As a result, it could be used on any type of network, including directed and undirected networks. Effective distance is an essential statistic in the networks G = (N, M) spreading process, which is employed in EDBC algorithm.

High influence neighbor nodes relate to automatic increase of the influence of the nodes. Furthermore, when the shortest distances between nodes rise, the impact of nodes on neighbors' declines. We can compute the effect of a node using the inverse square law as follows:

$$\left(I_{i,j}\right) = \sum_{j\varepsilon\eta(i)} \frac{\alpha(i,j) + d\left(v_i\right)}{\mathrm{ED}(i,j)^2}.$$

[28]

Where degree $d(v_i)$ denotes the degree of the node in terms of connected edges from it and $\alpha(i,j)$ is the power of K-shell as measurement of power influence.

The aggregate of a node's impact on all its neighbors, which measures that node's influence and can be computed as follows:

$$\mathrm{EDBC}(i) = \sum_{j\varepsilon\eta(i)} \left(I_{i,j}\right),$$

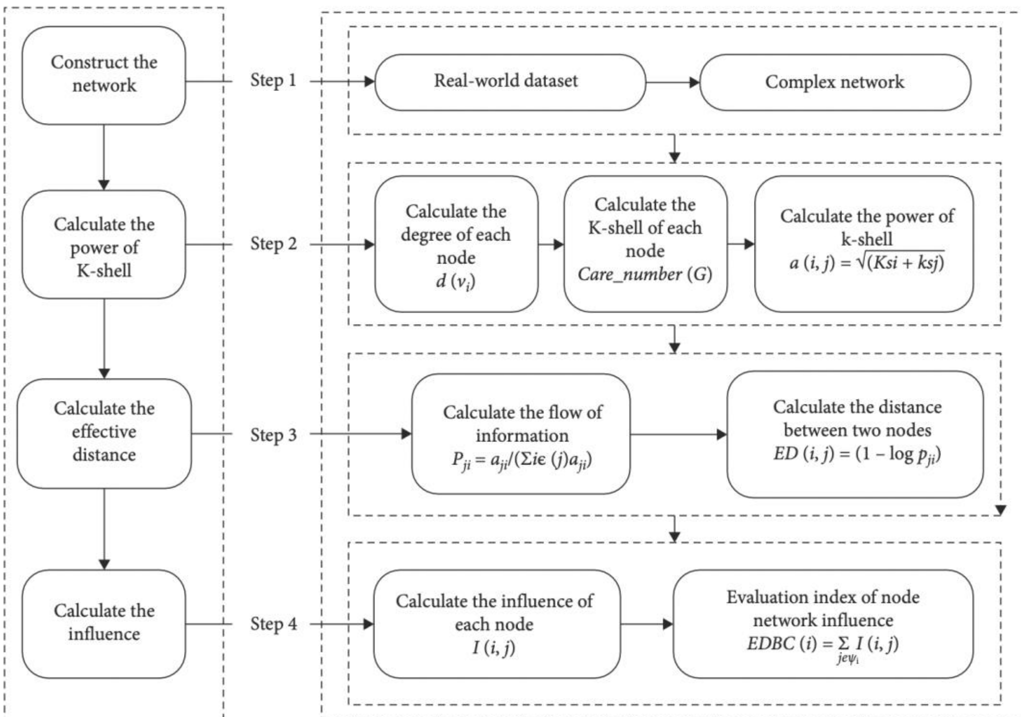
[28]

Three principles govern node influence in a network:

1. The node's position in the network; when a node is at the network's center, its effect is greatest. Otherwise, if the node is at the network's edge, its influence will be minimal.
2. The number of node neighbors: the more neighbors a node has, the more effect it has.
3. The distance between nodes: the shorter the distance, higher probability information will be transmitted between nodes and their neighbors.

The comprehensive mechanism of EDBC algorithm in [28] is displayed in the flowchart below. It is initialized with determination of the degree and K-shell of nodes in the network in the first step followed by measuring the power of K-shell. The distance between the nodes (nearest and next-nearest neighbors) is then calculated to render in the measurement of influence for each node.

A comparative study to analyze the efficiency of EDBC algorithm was evaluated with the help of Netscience study in [29] where an undirected and unweighted network represented co-authorships among scientists (as nodes) who were writing on the topic of networks. The results postulated that the key distinction between EDCC and conventional CC is that it uses effective distance instead of traditional geographic or binary distance. The EDCC contains more information about its neighbors than other centrality measures since it uses the node's input flow (in-degree in unweighted networks) and its neighbor's output flow (out-degree in unweighted networks). As a result, EDCC considers not only the network's global structure but also the local information of nodes. In comparison to conventional sorting algorithms and newly suggested various relevant algorithms, the experimental setup indicated that the proposed EDBC method is reasonable and substantial in terms of accuracy and effectiveness.



[29]

Fig 3: Flowchart of EDBC Algorithm Computation (Du Y, 2015)

3.7 Hierarchical Closeness

Hierarchical centrality paradigm offers valuable information about the underlying structure of complicated, real-world network topology that can't be completely realized by using traditional metrics. It is believed that hierarchical degree density is a logical extension of classical node degree density, capable of providing additional network hierarchy and connection information. The standard degree of a node in an augmented network, which includes all virtual connections generated during successive interactions, can also be viewed as the hierarchical degree of that node. A collection of acyclic subgraphs — the network's hierarchical backbone — may be extracted from any sophisticated directed network, providing significant data about its hierarchical structure. The paper [30] shows how construing the network weight matrix as a transition matrix allows the hierarchical backbone to be identified and characterized in terms of hierarchical degree, which expresses the total number of virtual edges established along successive transitions, and hierarchical successors, which expresses the number of nodes accessible from a specific node while moving up the scale.

Closeness Centrality, which measures how a node is close to all other nodes on average in a network, only provides centrality estimate from a global view. As a result, in real-world networks that are often made up of several communities connected, employing closeness centrality may have the drawback of neglecting the local central locations within communities. HCC (Hierarchical Closeness Centrality), on the other hand, proves to be a better index in finding the most influential vertices and community discovery inside networks as it is successful in depicting local centrality of vertices.

HCC evaluates the connectedness of a node with others that have equal or smaller closeness centrality by discarding the most central nodes in each tier of the network. Experiments on actual datasets reveals that HCC is more successful than GCC (Generalized Closeness Centrality) at recognizing influential vertices and locating communities. A parallel algorithm for computing HCC leveraging the observed adequate conditions for independently reconnecting vertices in a BFS tree was developed which results in reducing computational time. The discussion in [30] illustrates the comparison of two indexes – HCC and GCC - in the applications of maximum influence vertices identification and community detection. The modularity notion is a frequently used metric for characterizing the strength of division of a network into similar- behavior communities' identification as well as clustering. High modularity implies that nodes within communities have dense connections whereas nodes in segregated communities have sparse connections.

The above-mentioned principles and techniques for hierarchical properties of complex networks have been subjected to experimental data in word associations and zebrafish gene sequencing applications.

4. Conclusion

Centrality is a fundamental notion in the study of social and economic networks and different measures have been employed to quantify a node's importance, power, and prestige in a network. These different notions of centrality highlight the inherent different difficulty in selecting the optimal metric for a given network. The identification of influential nodes in a complex network is a challenging issue as there is ambiguity regarding the term 'influence'. Graph applications are notorious for exhibiting irregular access patterns and, hence it is necessary to establish axiomatic characterizations of centrality measurements in order to compare them. This provides a good understanding of the underlying computational structure of each centrality measure. along with provision of empirical evidence of correlation of centrality measures. The rigorous, well-founded assessment of closeness centrality measurements based on fundamental assumptions resulted in the emergence of seven closeness intrinsic methodologies. Random Walk centrality is a form of closeness centrality that quantifies the speed with which randomly walking messages reach a vertex from elsewhere in the graph. Hierarchical closeness (HC), an enhanced form of structural centrality, is an improved metric for ranking how centrally located a node is in a directed network. The hierarchical closeness explicitly includes information about the range of other nodes that can be affected by the given node. Individual 'Integration' refers to the degree an individual is connected to many and diverse others in a network. An integrated node maintains the indirect ties in addition with the first-order associations. An indicator of reachability, Radiality, describes the ability of an individual to form ties in the network to form novel paths for traversal of information within the network. Information centrality, measure of proximity, is concerned with the amount of time it takes traffic to reach the node (the information it receives), rather than the actual frequency of traffic passing through the node. The ease with which all other nodes in the network can access a node's functionality is quantified using the eccentricity metric. Effective Distance based centrality identifies influential nodes based on certain criterion using effective distance, K-shell, potential-like characteristics. It is true that often different centrality indices are designed to capture distinct aspects of the network under observation.

5. Future Work

In this paper, the focus is on providing holistic understanding of the extended measures based on centrality concept. The centrality measures discussion in this paper is restricted to the scope of undirected, simplex networks with consistent or uniform interaction within the nodes. I would like to explore and leverage efficient graph algorithms to cater to distributed, multiple layered networks induced with diverse relations. While a plethora of closeness related metrics have been considered in this analysis to study the underlying statistical characteristics and structural topography, a formal comparison of the seven metrics still needs to be devised based on a common application. Most of these measures have been independently assessed by the researchers which results to subjective perspective on the definition of the term ‘influence’.

My next line of work will focus on the investigation of performance evaluation of algorithms of closeness related metrics using similar parameters to recognize the impact of each on the network topology.

The current state-of-the-art in network centralities is based on a static or aggregated network representation. Real-world scenarios abound in dynamically evolving network topologies, where metrics can be applied to a range of system traces encompassing timestamp information, as well as dynamic networks that vary over time in general. I would like to assess the potential of these such modeling tools and metrics in other contexts.

Future avenues for research include enhanced granularity and heterogeneity on node and link breakage considering different vulnerabilities like increased network size and the applicability of different implementation methodologies.

References:

1. N. Meghanathan, "Use of Centrality Metrics to Determine Connected Dominating Sets for Real-World Network Graphs," 2015 12th International Conference on Information Technology - New Generations, 2015, pp. 243-248, doi: 10.1109/ITNG.2015.45.
2. A. Noori, "On the relation between centrality measures and consensus algorithms," 2011 International Conference on High Performance Computing & Simulation, 2011, pp. 225-232, doi: 10.1109/HPCSim.2011.5999828.
3. N. Meghanathan and R. Lawrence, "Centrality analysis of the United States network graph," 3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS 2016), 2016, pp. 1-6, doi: 10.1049/cp.2016.1499.
4. A. K. Jha and N. R. Sunitha, "Evaluation and optimization of smart cities using betweenness centrality," 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017, pp. 1-3, doi: 10.1109/ICAMMAET.2017.8186729.
5. M. T. Juzar and S. Akbar, "Buzzer Detection on Twitter Using Modified Eigenvector Centrality," 2018 5th International Conference on Data and Software Engineering (ICoDSE), 2018, pp. 1-5, doi: 10.1109/ICODSE.2018.8705788.
6. C. -C. Yen, M. -Y. Yeh and M. -S. Chen, "An Efficient Approach to Updating Closeness Centrality and Average Path Length in Dynamic Networks," 2013 IEEE 13th International Conference on Data Mining, 2013, pp. 867-876, doi: 10.1109/ICDM.2013.135.
7. Qin Qiong and Wang Dongxia, "Evaluation method for node importance in complex networks based on eccentricity of node," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), 2016, pp. 2499-2502, doi: 10.1109/CompComm.2016.7925149.
8. W. Li, M. Qiao, L. Qin, Y. Zhang, L. Chang and X. Lin, "Exacting Eccentricity for Small-World Networks," 2018 IEEE 34th International Conference on Data Engineering (ICDE), 2018, pp. 785-796, doi: 10.1109/ICDE.2018.00076.

9. K. Iwabuchi, G. Sanders, K. Henderson and R. Pearce, "Computing Exact Vertex Eccentricity on Massive-Scale Distributed Graphs," 2018 IEEE International Conference on Cluster Computing (CLUSTER), 2018, pp. 257-267, doi: 10.1109/CLUSTER.2018.00043.
10. H. Li, "Centrality analysis of online social network big data," 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), 2018, pp. 38-42, doi: 10.1109/ICBDA.2018.8367648.
11. Thomas W. Valente, Robert K. Foreman, “Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network”, Social Networks, Volume 20, Issue 1,1998, Pages 89-105
12. Raj, Mukund and Whitaker, Ross T, “Anisotropic Radial Layout for Visualizing Centrality and Structure in Graphs”, arXiv, 2017, doi: 10.48550/ARXIV.1709.00804
13. Ashtiani, M., Salehzadeh-Yazdi, A., Razaghi-Moghadam, Z. et al. A systematic survey of centrality measures for protein-protein interaction networks. BMC Syst Biol 12, 80, 2018, doi: 10.1186/s12918-018-0598-2
14. Zweig, K.A., et al.: Network analysis literacy. MMB & DFT 2014 p. 3 (2014)
15. Amrit, Chintan and ter Maat, Joanne, “Understanding Information Centrality Metric: A Simulation Approach”, arXiv, 2018, doi: 10.48550/ARXIV.1812.01292
16. Y. Ma, S. Wang and J. Tang, "Network Embedding with Centrality Information," 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017, pp. 1144-1145, doi: 10.1109/ICDMW.2017.162.
17. E. M. Daly and M. Haahr, "Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs," in IEEE Transactions on Mobile Computing, vol. 8, no. 5, pp. 606-621, May 2009, doi: 10.1109/TMC.2008.161.
18. K. Cai, H. Xie and J. C. S. Lui, "Information Spreading Forensics via Sequential Dependent Snapshots," in IEEE/ACM Transactions on Networking, vol. 26, no. 1, pp. 478-491, Feb. 2018, doi: 10.1109/TNET.2018.2791412.
19. Zhengyan Wu, Guiyan Jiang, Chunqin Zhang and Yong-yong Tang, "Traffic organization method for emergency evacuation based on information centrality," 2010 2nd International Conference on Advanced Computer Control, 2010, pp. 92-96, doi: 10.1109/ICACC.2010.5486749.

20. K. Dutta, M. Sharma, U. Sharma, S. K. Khatri and P. Johri, "Information Gain Model for Efficient Influential Node Identification in Social Networks," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 146-150, doi: 10.1109/AICAI.2019.8701344.
21. Landherr, A., Friedl, B. & Heidemann, J. A, 'Critical Review of Centrality Measures in Social Networks.', Bus *Inf Syst Eng* **2,** 371–385 (2010). https://doi.org/10.1007/s12599-010-0127-3
22. Hongjun Joo, Hung Soo Kim, Soojun Kim, Bellie Sivakumar, "Complex networks and integrated centrality measure to assess the importance of streamflow stations in a River basin",
Journal of Hydrology, Volume 598,2021, https://doi.org/10.1016/j.jhydrol.2021.126280
23. C. Mavroforakis, M. Mathioudakis and A. Gionis, "Absorbing Random-Walk Centrality: Theory and Algorithms," 2015 IEEE International Conference on Data Mining, 2015, pp. 901-906, doi: 10.1109/ICDM.2015.103
24. S. Simi and A. Sherin, "A centrality based random walk approach for topology formation in networks," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 1468-1472, doi: 10.1109/WiSPNET.2017.8300006
25. Alexandris, John and Stavrakakis, Gregory Karagiorgos 'and' Ioannis, "Enhanced Random Walk with Choice: An Empirical Study", 2010, arXiv, @misc{https://doi.org/10.48550/arxiv.1007.3157,doi:10.48550/ARXIV.1007.3157}
26. Weng, T., Zhang, J., Khajehnejad, M. *et al.* Navigation by anomalous random walks on complex networks. *Sci Rep* **6,** 37547 (2016), https://doi.org/10.1038/srep37547
27. Blöchl F, Theis FJ, Vega-Redondo F, and Fisher E: Vertex Centralities in Input-Output Networks Reveal the Structure of Modern Economies, Physical Review E, 83(4):046127, 2011
28. Ullah, Aman; Wang, Bin; Sheng, Jinfang; Long, Jun & Khan, Nasrullah, 2021, "Identification of Influential Nodes via Effective Distance-Based centrality Mechanism in Complex networks", 2021
29. Du Y, Gao C, Chen X, Hu Y, Sadiq R, Deng Y. A new closeness centrality measure via effective distance in complex networks. Chaos. 2015, doi: 10.1063/1.4916215

30. Jin, Hai, Chen Qian, Dongxiao Yu, Qiang-Sheng Hua, Xuanhua Shi, and Xia Xie. "Parallel computation of hierarchical closeness centrality and applications." World Wide Web 22.6 (2018): 3047-3064.