Descriptive Analysis of Social Network Measures

Himanshu Sharma, Vishal Srivastava

Abstract— **A large amount of work has been devoted to identifying community structures in social networks. A community is often described as a set of nodes that has more connections between its members than to the remainder of the network. The processes by which nodes come together, attract other neighboring nodes and develop communities over time is a central research issue in the social networks. Researchers find that the propensity of a node to join a network, and of networks to grow rapidly, depends in subtle ways on the underlying network structure. Often, networks have certain attributes that can be calculated to analyze the properties and characteristics of the network.**

In this paper, we characterize some network structure and network performance measures related to social networks. The analysis of these social network measures helps people to understand the formation of various types of social networks. From a network perspective, it is the structure of the network and how the structural properties affect behavior of a network that is informative, not simply the characteristics of the network members. The field of social network analysis is also focuses on the analysis of patterns of relationships among people, organizations, states and such social entities. SNA measures relationships among social actors, assesses factors that shape their structure, and ascertains the extent to which they affect social networks. So analysis of social network measures have significant importance in understanding the field of social network analysis, a field of great importance in today's and future scenario for extracting useful information from different types of social networks.

Index Terms— **Social network measures, Network properties, Social network analysis, network structure, network performance.**

I. Introduction

A social network mainly formed by joining of individuals sharing some common traits or having similar ideas about a subject. The tendency of people to come together and form groups is inherent in the structure of society; and the ways in which such groups take shape and evolve over time is a theme that runs through large parts of social science research [4]. The study of groups and communities is also fundamental in the mining and analysis of phenomena based on sociological data—for example, the evolution of informal close-knit groups within a large organization can provide insight into the organization's global decision-making behavior; the dynamics of certain subpopulations susceptible to a disease can be crucial in tracking the early stages of an epidemic; and the discussions within an Internet-based forum can be used to follow the emergence and popularity of new ideas and technologies.

Understanding the structure and dynamics of social networks is a natural goal for the field social network analysis. There are some social network properties and measures which helps in understanding the formation and evolution of social networks. These social network properties can be analyzed locally for a subgroup and globally for the whole network. These properties try to answer some important questions like what are the structural features that influence whether individuals will join a community, which communities will grow rapidly, How strongly or loosely the nodes are connected in a network, How fast will things move across the nodes in the network, Which node is most important in a network, Will conflicts most likely involve multiple groups or two factions, To what extent do the sub-groups and social structures overlap one another, etc. All of these aspects of sub-group and network structure can be very relevant to predicting the behavior of the network as a whole.

Despite the importance of such properties and their ability to detect the network behavior, less effort is made to measure these properties and understand the relationship among them and to other characteristics of social networks. This paper mainly focuses on the network structure and network performance measures and try to answer the many popular questions of the social network world.

II. Related Work on Social Network Measures

In the literature, a lot of work has been done on social networks, and analysis of social networks, but the researchers less discussed about the social network properties, and measures related to social networks. The effect of social network properties on a node, relationship of nodes, a group, its formation and growth is discussed by some researchers in their work. A.L. Barabási [20], provides an easy and readable introduction to the main models and properties of networks and their applications in many areas of real life, such as the spread of epidemics, fighting against terrorism, handling economic crises or solving social problems of the society. Blau [8], studied organizational and social structures, in particular bureaucracy. Many aspects of social phenomena, including upward mobility, occupational opportunity, and heterogeneity are discussed in his theories. He also produced theories on how population structures can influence human behavior. Blau's [16], contributes to social theory by work on exchange theory, which explains how small-scale social exchange directly relates to social structures at a societal level. He is the first one who map out the wide variety of social forces. Scott [2] provides an accessible introduction to the theory and practice of network analysis in the social sciences. It gives a clear and authoritative guide to the general framework of network analysis, explaining the basic concepts and technical measures, and reviewing the available computer programs. Scott [2] worked on issues of economic and

Himanshu Sharma, Computer Science & Engineering Department, Arya College of Engineering & Information Technology, Jaipur, Rajasthan, India, 9460867761.

Vishal Srivastava, Computer Science & Engineering Department, Arya College of Engineering & Information Technology, Jaipur, Rajasthan, India, 9214052387.

political sociology, social stratification, the history of sociology, and social network analysis. Scott [18], provides overview on concepts of centrality, sub-groups, adjacency and distance, which describe the quality of relationship ties and are visualized differently in sociograms. S. Wasserman and K. Faust [1], worked on social network analysis, its methods and applications, which focuses on relationships among social entities. They also used widely social network analysis in the social and behavioral sciences, as well as in economics, marketing, and industrial engineering. L. C. Freeman [5], discussed the concepts of point and graph centrality in social networks, also discussed and reviewed the different centrality measures and network measures of social networks. Freeman [11], discussed the structure and dynamics of social networks, and use graphic techniques for exploring social networks. P. Bonacich [13], discussed on centrality measures and the power and influence of the central node on other nodes and whole network. K. Faust [7], focused the work on local structure in social networks. Faust discussed the different local structures of the networks and their formation reasons and effects on whole network. T. Snijders [19], discussed about the degree and the degree variance of graph network. M. Granovetter [12], discussed about the strength of weak ties in his work. McPherson et.al. [22] has worked on homophily in social networks. Mustafa et.al. [29], discussed the impact of homophily on diffusion dynamics over social networks. R. Burt [30], describes the social structural theory of competition and discussed about various network performance measures like robustness, efficiency, diversity etc.

III. Present Work: Analysis of Social Network Measures

A social network most primarily can be analyzed on the basis of its representation models. Social networks generally represented by three methods: First is descriptive methods which use graphical representations, second is analysis methods which are based on adjacency matrix, and third is statistical models which are based on probability distributions. The most of the properties and measures of social networks are also based on these three representation methods of network data.

The other classification methods of network measures are as local measures, which analyze the network attributes with respect to individual units or dyads, and the global measures, which study the characteristics of the network considered as a whole. The two types of measures are not unrelated. Rather, the latter can be obtained from the former in a few instances by some sort of aggregation, as in the case of density or reciprocity.

According to S.Wasserman and K. Faust[1], The network data can be analyzed at various different levels These levels and their related measures are as follows:

Actor level: centrality, prestige and roles such as isolates, liaisons, bridges, etc.
Dyadic level: distance and reachability, structural and other notions of equivalence, and tendencies toward reciprocity.
Triadic level: balance and transitivity
Subset level: cliques, cohesive subgroups, components
Network level: connectedness, diameter, centralization, density, prestige, etc.

In this paper we seek to explore the social network measures by which groups develop and evolve in large-scale social networks and why these measures also indicates a lot of characteristics about a network like dense or sparse, weakly connected or strongly connected etc. There are mainly two types of network measures first one concerns about structure of networks and includes the important concepts of *density*, *centrality*, *betweenness* and *centralization*. Under these concepts are grouped several measures (or mathematical formulas) with various corresponding advantages and disadvantages regarding their use. Additionally, there are four measures of network performance: *robustness*, *efficiency*, *effectiveness* and *diversity*. This second set concerns the dynamics and thus depends on a theory explaining why certain agents do certain things (e.g., access to information).

IV. Different Measures of Social Networks

Network measures are classified in different related groups on the basis of some common parameters of analysis.

A. Measures Related to Networks And Actors

1) Size of a Network

The size of a network is often a very important. The size of a network can refer to the number of nodes or actors N or, less commonly, the number of edges E which can range from N-1 (a tree) to E_{max} (a complete graph). Now understand the importance of size of a network with some examples discussed here.

First, imagine a group of 12 students in a seminar. It would not be difficult for each of the students to know each of the others fairly well, and build up exchange relationships (e.g. sharing reading notes).

Now imagine a large lecture class of 300 students. It would be extremely difficult for any student to know all of the others, and it would be virtually impossible for there to be a single network for exchanging reading notes.

Size is critical for the structure of social relations because of the limited resources and capacities that each actor has for building and maintaining ties. As a group gets bigger, the proportion of all of the ties that could (logically) be present that is density will fall, and the more likely it is that why differentiated and partitioned groups will emerge.

In any network there are (N *N-1) unique ordered pairs of actors (that is AB is different from BA, and leaving aside self-ties), where N is the number of actors. So, a network of 10 actors, with directed data, there are 90 logically possible relationships. If we had undirected, or symmetric ties, the number would be 45, since the relationship AB would be the same as BA. The number of logically possible relationships then grows exponentially as the number of actors increases linearly. It follows from this that the range of logically possible social structures increases (or complexity increases) exponentially with size.

Fully saturated networks (i.e. one where all logically possible ties are actually present) are empirically rare, particularly where there are more than a few actors in the population.

2) *Density of a Network*

The density D of a network is defined as a ratio of the actual number of edges E to the number of maximum possible edges, given by D = 2E /N*(N-1). Another possible equation is D = T/N*(N-1) , whereas the ties T are unidirectional [1]. This gives a better overview over the network density, because unidirectional relationships can be measured.

Density is a number that varies between 0 and 1.0. When density is close to 1.0, the network is said to be dense, otherwise it is sparse. For quantitative data on relations, density could be defined as the mean strength of a tie. The problem with the measure of density is that it is sensible to the number of network nodes, therefore, it cannot be used for comparisons across networks that vary significantly in size [18].

As an example, consider the network named 33-physician influence network of Figure 1. In this directed influence network 163 nonzero ties were observed. Since 33*32=1056 ties were possible, network density is 0.154.

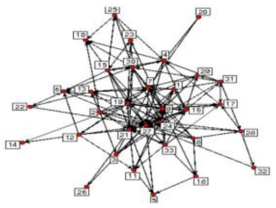


Fig. 1 Directed physician influence network

3) *Actor's Degree and the Degree Distribution*

In an undirected network, an actor's *degree* is the number of other actors to which it is directly connected. Analysis of directed networks distinguish between incoming and outgoing ties. The number of arcs oriented toward an actor is that actor's *in-degree* sometimes termed popularity or attractiveness; the number of arcs emanating from an actor is its *out-degree* also known as expansiveness.

Often actors having greater degrees have prominent roles in the network; indeed, the simplest measures of centrality are based on degree [5].

The *degree distribution* is the frequency distribution giving the number of actors having particular numerical degrees. Its variance measures the extent to which direct connectedness varies across actors Snijders [19].

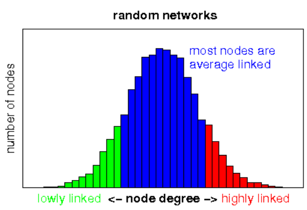


Fig. 2(a)

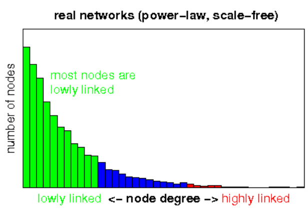


Fig. 2(b)

Fig 2. Shows node degree distribution curves for Random networks versus scale-free power-law networks

Figure 2(a) shows, that in traditional random networks most nodes have a medium node degree. The degrees of all nodes are distributed around the average. Figure 2(b) i.e. real networks often shows a skewed node-degree distribution in which most nodes have only few links but, by contrast, there exist some nodes which are extremely linked. This heavy tailed distribution is known as power-law or scale-free distribution. It should be noted that scale-free power-law node-degree distributions are not an universal characteristics of all real networks. It typically can be observed on sparsely connected networks. But more densely connected networks, by contrast, show an increasing divergence from power-law.

Barabási [20] and colleagues have focused on degree as their fundamental analytic interest Barabási [20], and Wolfram [21], showing that many network properties are shaped by the degree distribution.

As an example Figure 3 illustrate, networks with the same overall density but different degree distributions may have quite different structures. A "circle" network in which actor degree is constant (and hence, degree variance is 0), and a "star" network in which one actor has degree $N-1$ while all others have degree 1 lie at opposite ends of the spectrum with respect to degree variation.

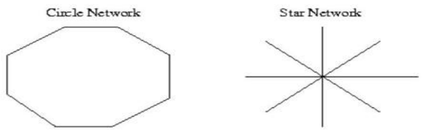


Fig. 3 Circle and star networks

B. *Measures Related to Social Distance And Related Concepts*

1) *Distance*

The properties of the network that we have examined so far deal primarily with immediate adjacency (the actor's immediate social neighbors). However, the connections of an actor's social neighbors can be very important, even if the actor is not directly connected to them (e.g. think of the importance of having "well-connected friends" in certain environments). In other words, sometimes being a "friend of a friend" may be quite consequential.

To capture this aspect of how individuals are embedded in networks, one approach is to examine *how far* (in terms of *social distance*) an actor is from others.

The *distance* between two actors is the minimum number of edges that takes to go from one to another actor. This is also known as the geodesic distance. If two actors are adjacent, the distance between them is 1 (i.e. it takes one step, or edge, to go from one to the other). If A links to B, and B links to C (and A does not link to C), then actors A and C are at a distance of 2, same like as in Figure 4.

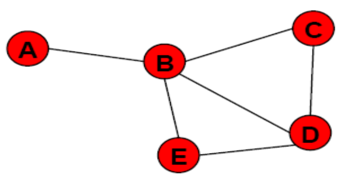


Fig. 4 An Undirected Graph

The distances among actors in a network may be an important macro-characteristic of the network as a whole. Where distances are great, it may take a long time for information to diffuse across a population. It may also be that some actors are quite unaware of, and influenced by others even if they are technically reachable, the costs may be too high to conduct exchanges. The variability across the actors in the distances that they have from other actors may be a basis for differentiation and even stratification. Those actors who are closer to more others may be able to exert more power than those who are more distant.

Sometimes we are also interested in studying the various ways that two actors, which are at a given distance, can be connected; multiple connections may indicate a stronger relation between two actors than a single connection**.**

2) Walks

The most general form of connection between two actors in a graph is called a walk. A walk is a sequence of actors and relations that begins and ends with actors. A closed walk is one where the beginning and end point of the walk are the same actor. Walks are unrestricted: a walk can involve the same actor or the same relation multiple times. The length of a walk is the number of edges that it uses.

Some examples of walks between A and C in the graph represented in Figure 4 are:

{A, B, C} length = 2; {A, B, D, C} length = 3;

{A, B, E, D, C} length = 4; {A, B, D, B, C} length = 4.

2) Cycles

A cycle is a specially restricted walk that is often used to examining the neighborhoods of actors (i.e. the points adjacent to a particular node). A cycle is a closed walk of 3 or more actors, all of whom are distinct, except for the origin / destination actor. There are no cycles beginning and ending with A in Figure 4, but there are 3 beginning and ending with actor B ({B, D, C, B}; {B, E, D, B}; {B, C, D, E, B}).

3) Trails

Sometimes it may be useful to study only those walks that do not re-use relations. A *trail* between two actors is any walk that includes any given relation at most once. (The same actors, however, can be part of a trail multiple times). The length of a trail is the number of relations in it. All trails are walks, but not all walks are trails. If the trail begins and ends with the same actor, it is called a closed trail.

In Figure 4, there are a number of trails from A to C. Excluded are tracings like {A, B, D, B, C} (which is a walk, but is not a trail because the relation BD is used more than once).

4) **Paths**

Perhaps the most useful definition of a connection between two actors (or between an actor and itself) is a *path*. A *path* is a walk in which each actor (and therefore each relation) in the graph may be used at most once. The single exception to this is a *closed path*, which begins and ends with the same actor. Paths may involve multiple adjacencies; the *length* of a path is the number of relationships or lines it contains.

In Figure 4, there are a limited number of paths connecting A and C: {A, B, C} Length=2; {A, B, D, C} Length=3; {A, B, E, D, C}Length=4.

For a network, matrix multiplication of an adjacency matrix **y** by itself yields the number of paths of a given length between any two actors. If no path from i to j exists, the geodesic distance from i to j is said to be infinite. In directed networks, the geodesic distance from i to j need not equal that from j to i.

5) Geodesic Distance

A *geodesic* path is a shortest-length path between a given pair of actors. *Geodesic distance* is defined as the length of a geodesic path and is perhaps the most widely-used network based measure of the social distance separating units/actors. The *length of a path* is the number of relations in it. The length of a shortest path between two actors is the geodesic distance between them. Thus, the geodesic distance between A and C in the graph of Figure 4 is 2.

If the network is *dense*, the geodesic path distances are generally small. This suggests that information may travel pretty quickly in this network. Also note that if there is a geodesic distance for each XY and YX pair, the graph is *fully connected*, and all actors are "*reachable*" from all others (that is, there exists a path of some length from each actor to each other actor). When a network is not fully connected, we cannot exactly define the geodesic distances among all pairs. The standard approach in such cases is to treat the geodesic distance between unconnected actors as a length greater than that of any real distance in the data like infinite. For each actor, we can calculate the mean and standard deviation of their geodesic distances to describe their closeness to all other actors.

For each actor, that actor's largest geodesic distance is called the *eccentricity* which is a measure of how far a actor is from the furthest other. Vertices with maximum eccentricity are called *peripheral vertices*. Vertices of minimum eccentricity form the *centre*.

we can calculate the mean (or median) geodesic distance and the standard deviation in geodesic distances for the matrix, and for each actor row-wise and column-wise. This would tell us how far each actor is from each other as a source of information for the other; and how far each actor is from each other actor who may be trying to influence them. It also tells us which actors behavior (in this case, whether they've heard something or not) is most predictable and least predictable.

5) Diameter of a network

To get another notion of the size of a network, we can think about its diameter. The *diameter of a network* is the largest geodesic distance in the (connected) network(if the network is not connected the largest distance is infinity). In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path lengths. The diameter of a network tells us how "big" it is, in one sense (that is, how many steps are necessary to get from one side of it to the other). The diameter is sometimes used as a measure of connectivity of a network.

The diameter is also a useful quantity in that it can be used to set an upper bound on the lengths of connections that we study. Many researchers limit their explorations of the connections among actors to involve connections that are no longer than the diameter of the network.

6) Flow, cohesion and influence

The use of geodesic paths to examine properties of the distances between individuals and for the whole network often makes a great deal of sense. However, there may be other cases where considering all connections among actors – not just the most efficient ones – may be more appropriate. *Rumours*, for instance, may spread in a network through all pathways – not just the most efficient ones. Similarly, how much credibility a person gives to a rumour may depend on how many times they hear it from different sources – not on how soon they hear it. For uses of distance like this, we need to take into account all of the connections among actors.

Thus, there are many networks where paths that are not necessarily the shortest will do the job (e.g. transmitting information, resources…) almost as well as shortest paths, particularly if the latter are problematic for some reason. It assumes that actors will use all pathways that connect them; the importance of each path can be weighted according to its length.

Several approaches have been developed which take into account all connections between pairs of actors. These measures have been used for a number of different purposes, and three are given below

Maximum flow

One notion of how totally connected two actors are (called maximum flow by UCINET) asks how many different actors in the neighborhood of a source lead to pathways to a target. If I need to send a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak - even if the person I send it to may have many ways of reaching you. If, on the other hand, there are four people to whom I can send my message, each of whom has one or more ways of retransmitting my message to you, then my connection is stronger. This "flow" approach suggests that the strength of my tie to you is no stronger than the weakest link in the chain of connections, where weakness means a lack of alternatives.

Hubbell and Katz cohesion

The maximum flow approach focuses on the vulnerability or redundancy of connection between pairs of actors - kind of a "strength of the weakest link" argument. As an alternative approach, we might want to consider the strength of all links as defining the connection. If we are interested in how much two actors may influence one another, or share a sense of common position, the full range of their connections should probably be considered.

Even if we want to include all connections between two actors, it may not make a great deal of sense (in most cases) to consider a path of length 10 as important as a path of length 1. The Hubbell and Katz approaches count the total connections between actors (ties for undirected data, both sending and receiving ties for directed data). Each connection, however, is given a weight, according to its length. The greater the length, the weaker the connection. How much weaker the connection becomes with increasing length depends on an "attenuation" factor. For example, if we have used an attenuation factor of .5. Then an adjacency receives a weight of one, a walk of length two receives a weight of .5, a connection of length three receives a weight of $(.5)^2$ i.e. (.25) etc.

Taylor's Influence

The Hubbell and Katz approach may make most sense when applied to symmetric data, because they pay no attention to the directions of connections (i.e. A's ties directed to B are just as important as B's ties to A in defining the distance or solidarity or closeness between them).

If we are more specifically interested in the influence of A on B in a directed graph, the Taylor influence approach provides an interesting alternative. The Taylor measure, like the others, uses all connections, and applies an attenuation factor. Rather than standardizing on the whole resulting matrix, however, a different approach is adopted. The column marginals for each actor are subtracted from the row marginals, and the result is then normed. Here we look at the balance between each actor's sending connections (row marginals) and their receiving connections (column marginals). Positive values then reflect a preponderance of sending over receiving to the other actor of the pair -- or a balance of influence between the two.

C. Measures Related to Connections And Connectivity

1) Connectedness

Connectivity is a property of a network (not of its individual actors) that extends the concept of adjacency. If it is possible to establish a path from any actor to any other actor of a network (e.g. every actor is reachable by every other one), the network is said to be *connected*; otherwise the network is *disconnected*. The way in which a network is connected plays a large part into how networks are analyzed and interpreted. Networks are classified in four different categories on the basis of it connectedness:

Clique/Complete Graph: a completely connected network, where all nodes are connected to every other node. These networks are symmetric in that all nodes have in-links and out-links from all others.

Giant Component: A single connected component which contains most of the nodes in the network.

Weakly Connected Component: A collection of nodes in which there exists a path from any node to any other, ignoring directionality of the edges.

Strongly Connected Component: A collection of nodes in which there exists a *directed* path from any node to any other.

2) Reachability

An actor is "reachable" by another if there exists any set of connections by which we can trace from the source to the target actor, regardless of how many others fall between them. If the data are asymmetric or directed, it is possible that actor A can reach actor B, but that actor B cannot reach actor A. With symmetric or undirected data, of course, each pair of actors either are or are not reachable to one another. If some actors in a network cannot reach others, there is the potential of a division of the network. Or, it may indicate that the population we are studying is really composed of more than one separate sub-population.

D. Measures Related to Local Structures in Networks

So far we have looked mainly the ways in which individuals are connected, and the distances between them. In this section we look at the same issue of connection but this time our focus is the social structure, rather than the individual: here we adopt a slightly more “macro” perspective that focuses on the local structures within which individual actors are embedded.

1) The dyad census and reciprocity

The smallest social structure in which an individual can be embedded is a *dyad* (i.e. a pair of actors). In binary-valued directed networks as in Figure 5, three types of dyadic relationships may exist: *mutual* dyads, in which a tie from *i* to *j* is accompanied by one from *j* to *i; asymmetric* dyads in which there is a relationship between *i* and *j* in one direction, but not the other; and *null* dyads in which there is no tie in either direction. The *dyad census* is the set of three network statistics giving the number of each dyad type found within a given network.

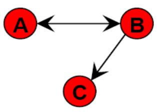


Fig. 5 A directed graph showing 3 types of dyadic relations

If all ties in a binary network are either mutual or null, the network is said to be *symmetric*, in which case the adjacency matrix $\mathbf{y}$ and its transpose $\mathbf{y}^T$ are identical; an undirected network is symmetric by construction. The presence and magnitude of a tendency toward symmetry or reciprocity in a directed network can be measured by comparing the number of mutual dyads to the number expected under a model in which ties are reciprocated at random. If the number of mutuals is lower than expected, there is a tendency away from reciprocation.

A network that has a predominance of null or reciprocated ties over asymmetric connections may be more “equal” or “stable” than one with a predominance of asymmetric connections, So reciprocity in a social network indicates some sort of balance or harmony, which can nullify the negative effects of social stratification. A potentially interesting analysis is to study the extent to which a population is characterized by reciprocated ties; this may tell us about the degree of cohesion, trust, and social capital that is present.

Local reciprocity of a vertex in the network of a social choice relation is an indicator of its social congeniality or level of being integrated with others in the network. Global reciprocity, on the other hand, is a measure of integration of its vertices among themselves. Hence, it becomes a measure of social solidarity of a group or community.

2) The Triad Census, Transitivity, and Closure

Triads in undirected, binary networks may include four possible types of triadic relations (no ties, one tie, two ties, or all three ties). Triads having 3 relationships are said to be *closed* or *transitive*, in that each pair of units/actors linked by a direct tie is also linked by an indirect path through the third unit/actor. Counts of the relative prevalence of these four types of relations across all possible triples can give a good sense of the extent to which a population is characterized by “isolation,” “couples only,” “structural holes” (i.e. where one actor is connected to two others, who are not connected to each other), or “clusters”.

For directed binary networks, 16 distinct triad types exist, distinguished by the number and orientation of the directed ties they include [1]. These triadic relationships suggests about the hierarchy, equality, and the formation of exclusive groups in a network. To identify the frequency of each of these relations we may wish to conduct a “triad census” for each actor, and for the network as a whole. In particular, we may be interested in the proportion of triads that are “transitive” (i.e. those that display a type of balance where, if A directs a tie to B, and B directs a tie to C, then A also directs a tie to C). Triad types which includes transitive substructures are indicative of the network closure.

3) Cliques

A clique is a subset of the vertices such that every pair of vertices in the subset is connected by an edge. Figure 6 shows example of a clique.

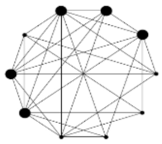


Fig. 6 Example of a Clique

4) N-Cliques

The strict definition of clique (i.e. everyone is connected to everyone else) may be too strong for some purposes. A more general approach is to define an actor as a member of a clique if it is connected to every other member of the clique at a distance no greater than a given number. This approach to defining sub-structures is called N-clique, where N stands for the length of the path allowed to make a connection to all other members.

5) N-Clans

An N-clan is an N-clique where all ties among members of the N-clique occur through members of the N-clique.

6) Clustering coefficient

In many large networks, a very large proportion of the total number of ties are highly “clustered” into local neighbour -hoods. To be precise, the density in local neighbourhoods of many large networks tends to be much higher than we would expect for a random graph of the same size.

The clustering coefficient is a measure of an "all-my-friends-know-each-other" property. This is sometimes described as the friends of my friends are my friends. More precisely, the clustering coefficient of a node is the ratio of existing links connecting a node's neighbors to each other to the maximum possible number of such links. For nodes with fewer than two neighbours the clustering coefficient is undefined

The clustering coefficient for the entire network is the average of the clustering coefficients of all the nodes. A high clustering coefficient for a network is another indication of a small world. Some analysts use a “weighted” version of the clustering coefficient, giving a weight to the neighbourhood densities proportional to their size (i.e. actors with larger neighbourhoods get more weight in computing the average density).

For example consider Figure 7, the clustering coefficient of a node A (blue color) is 1 if every neighbour connected to A is also connected to every other node within the neighbourhood of A, and 0 if no node that is connected to A connects to any other node that is connected to A.

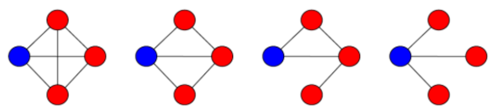


Fig. 7 Clustering coefficient of the blue node in various undirected networks (from left to right): 3/3, 2/3, 1/1, 0/3.

So, the clustering coefficient of the *i*'th node is

$$C_i = \frac{2e_i}{k_i(k_i - 1)},$$

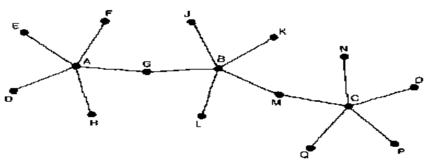
where k_i is the number of neighbours of the *i*'th node, and e_i is the number of connections between these neighbours.

D. Measures Related to Centrality and Centralization

Measures of centrality reflect the prominence of actors/ units within a network and identify the most important nodes within a network. An actor's prominence reflects its greater visibility to the other network actors. Measures of centrality are among the most widely-used actor-level measures that derive from network data. The concept of centrality encompasses two levels: *Local* and *Global.*

Local centrality

Local centrality measures are expressed in terms of the number of nodes to which a node is connected directly. (only considers direct ties means the ties directly connected to that node). A node is locally central when it has the higher number of ties with other nodes. The local centrality measure is degree centrality. Figure 8 shows the absolute and relative local centrality of some points of the given graph.



	A, C	B	G,M	J,K,L	All Other Points
Local Centrality – Absolute	5	5	2	1	1
Local Centrality – Relative	0.33	0.33	0.13	0.07	0.07
Global Centrality	43	33	37	48	57

Fig. 8 Showing Local and Global Centrality of different points.

Global centrality

Global centrality considers both direct and indirect ties (which are not directly connected to that node). *Global centrality* is expressed in terms of the *distances* among the various nodes. Figure 8 shows the global centrality of some points of the given graph.

A node is globally central if it lies at short distance from many other nodes. Such node is said to be "close" to many of the other nodes in the network, sometimes global centrality is also called *closeness*. The global centrality measures include closeness centrality, betweenness centrality, eigen vector centrality (Bonacich power centrality).

1) Degree centrality

Historically first and conceptually simplest is *degree centrality*, which is defined as the number of links incident upon a node (i.e., the number of ties that a node has). In undirected data, actors differ from one another only in how many connections they have. In the case of a directed network (where ties have direction), we usually define two separate measures of degree centrality, namely indegree and outdegree.

Accordingly, *indegree* is a count of the number of ties directed to the node/actor. If an actor receives many ties, they are often said to be *prominent*, or to have high *prestige*. That is, many other actors seek to direct ties to them, and this may indicate their importance. Figure 9(a) shows the indegree of node X.

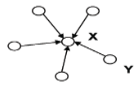


Fig. 9(a) indegree of node X=5

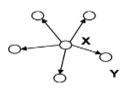


Fig. 9(b) outdegree of node X=5

Fig. 9 Showing indegree and outdegree of node X.

Outdegree is the number of ties that the node/actor directs to others. Actors who have unusually high out-degree are actors who are able to exchange with many others, or make many others aware of their views. Actors who display high out-degree centrality are often said to be *influential* actors. Figure 9(b) shows the outdegree of node X.

Actors who have more ties to other actors may be in advantageous position. Because they have many ties, they may have alternative ways to satisfy needs, and hence are less dependent on other individuals. Because they have many ties, they may have access to, and be able to call on more of the resources of the network as a whole. Because they have many ties, they are often third parties and deal makers in exchanges among others, and are able to benefit from this brokerage. So, a very simple, but often very effective measure of an actor's centrality and *power potential* is their degree.

The degree centrality index C_D for node n_i is given as

$$\mathbf{C_D(n_i) = d_i(n_i)}$$

where d_i is degree of node n_i.

To standardize or normalize the degree centrality index, so that networks of different sizes (g) may be compared, divide degree centrality by the maximum possible indegrees (= g-1 nodes if everyone is directly connected to i), and express the result as either a proportion or percentage:

$$\mathbf{C'_D(n_i) = C_D(n_i) / (g-1)}$$

As example see Figure 10 , here in star graph, the most central actor (n_1) has degree centrality = 6 but the six peripheral actors each have degree centrality = 1; their standardized values are 1.00 and 0.167, respectively.

All seven circle graph actors have identical degree centrality (=2), so no central actor exists; their standardized values are each 0.333.

In the line graph, the two end actors have smaller degree centralities (degrees = 1) than those in the middle (=2); the respective standardized scores are 0.167 and 0.333.

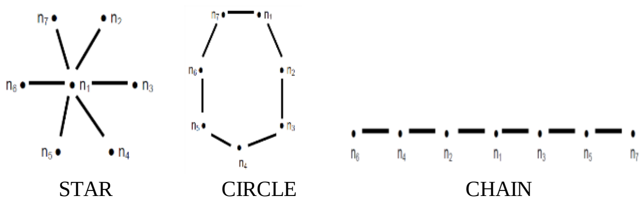


Fig. 10 Different types of Examples for explaining degree centrality

2) Closeness centrality

Degree centrality measures might be criticized because they only take into account the immediate ties that an actor has, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighborhood.

Closeness centrality approaches emphasize the distance of an actor to all others in the network by focusing on the geodesic distance from each actor to all others. One can consider either directed or undirected geodesic distances among actors. The sum of these geodesic distances for each actor is the "*farness*" of the actor from all others. We can convert this into a measure of nearness or closeness centrality by taking the reciprocal (that is one divided by the farness) and norming it relative to the most central actor. Thus, the more central a node is the lower its total distance from all other nodes.

An actor that is close to many others can quickly interact and communicate with them without going through many intermediaries. Thus, if two actors are not directly tied, requiring only a small number of steps to reach one another is important to attain higher closeness centrality.

Actor closeness centrality is the inverse of the sum of geodesic distances from actor i to the g-1 other actors (i.e., the reciprocal of its "farness" score):

$$\mathbf{C_C(n_i)} = \left[\sum_{\mathbf{j=1}}^{\mathbf{g}} \mathbf{d(n_i, n_j)}\right]^{-1}$$

Closeness can be calculated only for a connected graph, because distance is "infinite" (undefined) if members of a nodal pair are not mutually reachable (no paths exist between i and j). To standardize or normalize the closeness centrality index divide it by a maximum possible distance, and express the result as either a proportion or percentage:

$$\mathbf{C'_c(n_i) = C_c(n_i) / (g\text{-}1)}$$

As example see Figure 10, Here in the star graph, actor n_1 has closeness = 1.0 while the six peripheral actors = 0.545.
All circle graph actors have the same closeness (0.50).
In the chain graph, the two end actors are less close (0.286) than those in the middle (0.50).

3) Betweenness centrality

Betweenness is a centrality measure of a node within a network. Betweenness measures the extent to which a particular node lies "between" the various other nodes in the network. Betweenness centrality quantifies the number of times a node acts as a *bridge* along the shortest path between two other nodes. It was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network by Linton Freeman[18]. In his conception, vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness.

Betweenness centrality views an actor as being in a favored position to the extent that the actor falls on the geodesic paths between other pairs of actors in the network. That is, the more people depend on me to make connections with other people, the more power I have. If, however, two actors are connected by more than one geodesic path, and I am not on all of them, I lose some power. The betweenness of a node measures the extent to which an agent (represented by a node) can play the part of a broker or gatekeeper with a potential for control over others. As a cutpoint in the shortest path connecting two other nodes, a between actor might control the flow of information or the exchange of resources, perhaps charging a fee or brokerage commission for transaction services rendered.

Actor betweenness centrality for actor i is the sum of the proportions, for all pairs of actors j and k, in which actor i is involved in a pair's geodesic(s)

$$\mathbf{C_B(n_i)} = \sum_{\mathbf{j<k}} \frac{\mathbf{g_{jk}(n_i)}}{\mathbf{g_{jk}}}$$

where $\mathbf{g_{jk}}$ is total number of geodesic paths from node *j* to node *k* and $\mathbf{g_{jk}(n_i)}$ is the number of those geodesic paths that pass through $\mathbf{n_i}$.

As with the other centrality standardizations, normalize the betweenness centrality scores by dividing them by the maximum possible betweenness, expressed as proportion or percentage. The betweenness may be normalized by dividing it through the number of pairs of vertices not including $\mathbf{n_i}$, which for directed graphs is *(n-1)(n-2)* and for undirected graphs is *(n-1)(n-2)/ 2* .

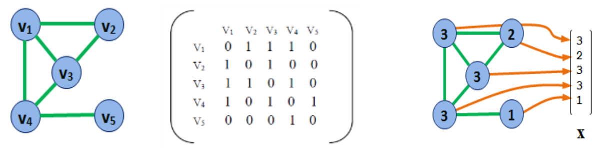
Consider Figure 10 as an example, In the star graph, actor n_1 has betweenness = 1.0 while the six peripheral actors= 0.0. All circle graph actors have the same betweenness (0.2).
In the chain graph, the two end actors have no betweenness (0.0), the exactly middle actor n_1 has the highest betweenness (0.60), while the two adjacent to it are only slightly less central (0.53).

4) Eigenvector Centrality

The eigenvector approach is an effort to find the most central actors (i.e. those with the smallest farness from others) in terms of the "global" or "overall" structure of the network, and to pay less attention to patterns that are more "local." . Others approaches include Katz centrality, Bonacich's power centrality, and Google Page Rank centrality.
Eigenvector centrality is one of several node metrics that characterize the "global" (as opposed to "local") prominence of a vertex in a graph. The abstract of eigenvector centrality is to compute the centrality of a node as a function of the

centralities of its neighbors. Eigenvector centrality also brings the graph in equilibrium position. Eigenvector centrality is also a measure of the *influence* of a node in a network

To Understand it better take an example. Consider the graph below and its 5x5 adjacency matrix, A.



Matrix - A

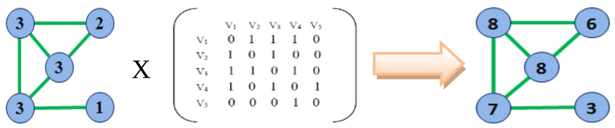
Fig. 11(a) Fig. 11(b) Fig. 11(c)

Fig. 11 (a) Shows a undirected graph, (b) its adjacency matrix A, (c) degree of each vertex and its matrix X.

And then consider, X, a 5x1 vector of values, one for each vertex in the graph. In this case, we've used the degree centrality of each vertex. Now when we multiply the vector X by the matrix A. The result, of course, is another 5x1 vector.

$$A \times X = \begin{bmatrix} 8 \\ 6 \\ 8 \\ 7 \\ 3 \end{bmatrix}$$

If we look closely at the first element of the resulting vector we see that the 1s in the A matrix "pick up" the values of each vertex to which the first vertex is connected (in this case, the second, third, and fourth) and the resulting value is the sum of the values each of these vertices had. In other words, what multiplication by the adjacency matrix does, is reassign each vertex the sum of the values of its neighbor vertices.



Matrix - A

Fig. 12 Shows the spread out effect of degree centrality

This has, in effect, "spread out" the degree centrality. That this is moving in the direction of a reasonable metric for centrality can be seen better if we rearrange the graph a little bit as shown in Figure 13.

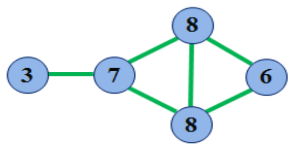


Fig. 13 Rearrange of Graph of Figure 12 to understand the spread out effect of degree centrality

Suppose we multiplied the resulting vector by A again, how might we interpret what that meant? In effect, we'd be allowing this centrality value to once again "spread" across the edges of the graph. And we'd notice that the spread is in both directions (vertices both give to and get from their neighbors). We might speculate that this process might eventually reach an equilibrium when the amount coming into a given vertex would be in balance with the amount going out to its neighbors. Since we are just adding things up, the numbers would keep getting bigger, but we could reach a point where the share of the total at each node would remain stable.

At that point we might imagine that all of the "centrality-ness" of the graph had equilibrated and the value of each node completely captured the centrality of all of its neighbors, all the way out to the edges of the graph.

The vector notation equation for finding eigenvectors centrality is:

AX = λ X (Eigen Vectors)

In general, there will be many different eigen-values λ for which an eigen-vector solution exists.

It is important to remember that centrality indices are only accurate for identifying the most central nodes. The measures are seldom, if ever, meaningful for the remainder of network nodes. Also, their indications are only accurate within their assumed context for importance, and tend to "get it wrong" for other contexts.

Centralization

Centralization provides a measure on the extent to which a whole network has a centralized structure (as shown in Figure 14). Whereas density describes the general level of connectedness in a network; centralization describes the extent to which this connectedness is organized around particular focal nodes. Centralization and density, therefore, are important complementary measures. The general procedure involved in any measure of network centralization is to look at the differences between centrality scores of the most central node and those of all other nodes. Centralization is then the ratio of the actual sum of differences to the maximum possible sum of differences [18].

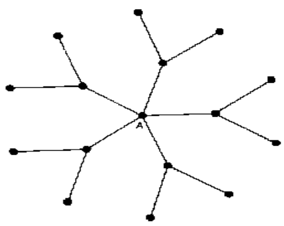


Fig. 14 A highly centralized graph

There are three types of graph centralization – one for each of the 3 centrality measures: local, global and betweenness. All 3 centralization measures vary from 0 to 1.0. A value of 1.0 is achieved on all 3 measures for "star" networks.

0 corresponds to a network in which all the nodes are connected to all other nodes. Between these two extremes lie the majority of the real networks. Methodologically, the choices of one of these 3 centralization measures depend on which specific structural features the researcher wants to illuminate.

For example, a local centrality based measure of network centralization seems to be particularly sensitive to the local dominance of nodes, while a betweenness-based measure is rather more sensitive to the chaining of nodes.

E. Measures Related to Homophily

Homophily (i.e., "love of the same") is the tendency of individuals to associate and bond with similar others. Many studies that have observed homophily in some form or another and they establish that similarity breeds connection [22]. These include age, gender, class, and organizational role. Individuals in homophilic relationships share common characteristics (beliefs, values, education, etc.) that make communication and relationship formation easier. Homophily is a metric studied in the field of social network analysis in which it is also known as assortativity. Homophily between mated pairs in animals has been extensively studied in the field of evolutionary biology in which it is known as assortative mating.

Types of Homophily

Lazarsfeld and Merton [23] distinguished homophily as *status homophily* and *value homophily*. It is simply the amount of homophily that would be expected by chance [22].

1) Status homophily

Individuals with similar social status characteristics are more likely to associate with each other than by chance, such as ascribed characteristics like race, ethnicity, sex, age, and acquired characteristics like religion, education etc [22][23].

2) Value homophily

In value homophily individuals tends to associate with others who think in similar ways, regardless of differences in status [22][23].

Many other Researchers have distinguished homophily as *baseline homophily* and *inbreeding homophily*. It is the amount of homophily over and above the "expected by chance" value [22].

Dimensions of Baseline homophily and Inbreeding homophily

1) Race and Ethnicity

Race and Ethnicity is of the most importance in dividing social network in United States today. Strong structural effects of category size and category differences impact largely on Race and Ethnicity. For Race and Ethnicity dimension, the baseline homophily generate by different group sizes usually combines with the different racial/ethnic groups position on other aspects like education, occupation, and etc. The baseline homophily plays an important role not only in large population, but also in smaller ones like classroom and workplaces. In addition, beside the baseline homophily, the primary standing of racial/ethnic homophily also accounts to the greatest proportion of inbreeding homophily [22].

2) Sex and Gender

With regard to Sex and Gender, homophily of networks is remarkably opposite to that of Race and Ethnicity. Compared with Race and Ethnicity, men and women link with each other with considerable connections in residence, social class and other characteristics. In addition they are considered roughly equal in population. It is concluded that inbreeding but not baseline homophily leads to most Sex homophily [22].

3) Age

Age homophily is resulted from the baseline homophily in a high extend. An interesting pattern of Age homophily for groups with different ages was found by Marsden [24]. It was indicated a strong relationship between someone's age and the social distance to other people with regard to confiding in someone. For example, the larger age gap someone had, the smaller chances that they were confided by others with lower ages to "discuss important matters" [22].

5) Religion

Both Baseline homophily and Inbreeding homophily causes the dimension of Religion [22].

6) Education, Occupation and Social Class

The dimensions of these homophily were claimed to be derived greatly from one's family of origin. It indicated that regarding Education, Occupation and Social Class, the Baseline homophily account for a large proportion of them [22].

Causes of Homophily

1) Geography

The most basic source of homophily is space. People are more likely to have contact with those who are closer to us in geographic location than those who are distant. Though the telephone, e-mail, social network have loosened the bounds of geography by lowering the effort involved in contact, these new modes have certainly not eliminated the old pattern.

2) Family ties

Family connections, of strong affective bonds and slow decay, are the bio-social web that connect us to those who are simultaneously similar and different. Kin ties often produce relatively close, frequent contact among those who are at great geographic distance. On the other hand, the marriage bond within families creates rather dramatic structuring of kinship ties.

3) Organizational Foci

School, work and voluntary organizational foci provide the great majority of ties that are not kin, supporting the argument that focused activity puts people into contact with one another to foster the formation of personal relationships. Many friendships, confiding relations, and social support ties are formed within voluntary groups. The social homogeneity of most organizational foci creates a strong baseline homophily in networks that are formed there.

4) Isomorphic Sources: Occupational, Family, and Informal roles

The connections between people who occupy equivalent roles will induce homophily in the system of network ties, which is common in three domains: workplace, family, and informal networks.

5) Cognitive processes

People who have demographic similarity tend to own shared knowledge, and therefore they have a greater ease of communication and share cultural tastes, which can also generate homophily.

Impact of homophily

Homophily facilitates individual's social interactions. For example: homophily is regarded as an explanation for the appearance of some qualities such as being tolerant, cooperative are localized in social space [25]. Also, homophily helps people to access information [26], diffusion of innovations and behaviors [27], opinion and norm formation [28]. Homophily often leads to homogamy that is marriage between people with similar characteristics [22]. Homophily influences diffusion patterns over a social network via two approaches. First, homophily has an impact on the way a social network develops. Second, individuals are easier to make social influence on people alike [29].

F. Measures related to Network performance

Network's performance can be evaluated as a combination of (1) its robustness to the removal of ties and/or nodes. (2) Its efficiency in terms of the distance to traverse from one node to another and its non-redundant size. (3) Its effectiveness in terms of information benefits allocated to central nodes and (4) its diversity in terms of the history of each of the nodes.

1) Robustness

Social network analysts have highlighted the importance of network structure in discussion of network's *robustness*. The robustness can be evaluated by studying how it becomes fragmented as an increasing fraction of nodes is removed. Robustness is measured by an estimate of the tendency of individuals in networks to form local groups or *clusters* of individuals with whom they share similar characteristics, i.e., *clustering*. E.g., if individuals A, B, and C are all bioinformatics experts and if A knows B and B knows C, then it is highly likely that A knows C. When the measure of the clustering of individuals is high for a given network, the robustness of that network increase – within a cluster/group where everyone knows everybody it is unlikely that a given person will serve as a lynchpin in the network, potentially destroying connectivity within the network by leaving.

2) Efficiency

Efficient networks are those in which nodes (individuals or firms) can access instantly a large number of different nodes – sources of knowledge, status, etc., through a relatively small number of ties, Burt [30] call these nodes *non- redundant contacts*. Given two networks of equal size, the one with more non-redundant contacts provides more benefits. There is little gain from a new contact redundant with existing contacts. Time and energy would be better spent cultivating a new contact to un-reached people Burt [30]. Social network analysts measure efficiency by the number of non-redundant contacts and the average number of ties an ego has to traverse to reach any alter, this number is
referred to as the *average path length*. The shorter the average path length relative to the size of the network and the lower the number of redundant contacts and the more efficient is the network.

3) Effectiveness

While efficiency targets the reduction of the time and energy spent on redundant contacts by, e.g., decreasing the number of ties with redundant contacts, *effectiveness* targets the cluster of nodes that can be reached through non-redundant contacts. Each cluster of contacts is an independent source of information. According to Burt [30], one cluster around this non-redundant node, no matter how numerous its members are, is only one source of information, because people connected to one another tend to know about the same things at about the same time. For example, a network is more effective when the information benefit provided by multiple clusters of contacts is broader, providing better assurance that the central node will be informed. Moreover, because non-redundant contacts are only connected through the central node, the central node is assured of being the first to see new opportunities created by needs in one group that could be served by skills in another group Burt [30].

4) Diversity

While efficiency is about reaching a large number of (non-redundant) nodes, node's diversity, suggests that it is critical from a performance point of view that those nodes are diverse in nature i.e. the history of each individual node within the network is important. If in a graph a node has many non redundant paths to reach the target node then the node or graph is said to be diverse in nature. Figure 15 shows diversity of a graph before and after removing some edges. As non-redundant paths decreases diversity of a graph also decreases.

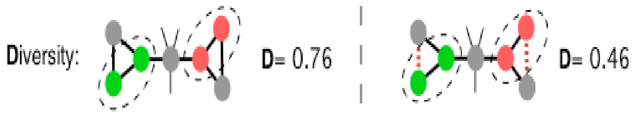


Fig. 15(a) Diversity of whole graph

Fig. 15(b) Diversity of graph after deleting some edges(dotted lines)

Fig. 15 Showing diversity of a graph

V. Comparison Between Various Social Network Measures

In this section, we compared the various social network measures discussed in this paper and try to find out the similarities and dissimilarities between them. Table I shows the comparison of various social network measures, what they indicate and their effect on neighboring nodes, the group and the whole network. This summarized data help us to find the various interdependent and related measures. This table also gives the answer of some important questions of social network field like how a node becomes a leader in group or network and controls the flow of information, Why some nodes are more informed than other, Why some intermediate nodes are important and acts as a broker,Why some resources do not reach some places or prevented from getting them, Why some relations are strong even after the large distances etc.

Social Network Measures	Value	Indication	Effect / Influence
Size of a Network	Increase	-Indicates density fall. -Complexity of network increases.	-Differentiated/Partitioned groups emerge. -Effects limited resources & relationships.
Density of a Network	High	-Well connected network - Dense network -Strength of ties/relationship	-Closeness increases
Degree	Increase	-Indegree indicates popularity, prominence & prestige of a node. -Outdegree indicates the influence power of a node. -Having more opportunities and alternatives -Less dependent on other nodes, hence powerful. -a node having high degree means more prominent role in network	-Connectedness of network - role of a node -Power of a node - Controls flow of information
Degree Distribution	-	-Probability distribution of degree -Shows proportion of nodes having high, average & low degree in network. -Important nodes in graph	- Structure of network -Prominence of nodes
Distance	Increase	- farness of node increases. -Reachability of a node -Power of node decreases - Influence on other actors decreases -Information diffusion time increases	- Size of network -Closeness of a node - May be possible that resources either do not reach to distant nodes or cost to reach them is high -time to reach
Paths	High number	- chances to Reach a node increases -connectedness of graph -Number of alternative ways	-Influence of a node.
Geodesic Distance	Low	-Shortest length path between 2 actors. -Dense network normally - Connected graph or not - Reachability of a node	-Influence of a node -time to reach -closeness of node
Diameter of a network	Increase	- Largest geodesic distance -Size of a network increases -Measure of connectivity	- A disconnected graph has infinite diameter, i.e. new group formed.
Maximum flow	-	-Strength of the node is equals to strength of weakest tie -Relative importance of a node	- Number of alternatives.
Cohesion	High	- Take all links into account -Indicate dense network -Closeness between nodes -Depends on distance	-Position of a node -Influence on a node
Influence	High	-Influence on other nodes -Closeness between nodes -Power of a node	-Closeness between nodes. - Controls flow of information
Connectedness	High	-Graph is connected or not - Reachability	- Community Formation -Density of graph
Reachability	-	- Reach to a node -If any part of graph is not reachable, it means it consists of more than one community. - Influence on other nodes	-Community Formation -Resource getting -Awareness or knowledge
The dyad census and reciprocity	High	-Balance or harmony in network -Cohesion in a network -Trust in a network - Solidarity of a group -Strength of relationship	-Balance of network -Strength of relationship
The Triad Census, Transitivity, and Closure	High	- Hierarchy, equality in a group -Network closure -Group formation	-Strength of relationship -Group formation
Clustering coefficient	High	- Measure of an "all-my- friends-know-each-other" property. - indication of small world network -High tendency to form cluster - Short distances	-Cluster formation -Geographically short distances.
Local centrality	High	-Degree of a node is high -Power of a node locally -Short distances -importance of a node	-Importance of a node -Power of a node locally -Information flow -Awareness of node
Global centrality	High	-Closeness -Betweenness -Power of a node globally -Prominent Role	-Importance of a node -Power of a node globally -Information flow -Awareness of node
Degree centrality	High	-Indegree indicates prominence & prestige of a node. -Outdegree indicates the influence of a node on others. -Having more oppurtunities and alternatives -Less dependent on other nodes, hence powerful.	-Role of a node -Power of a node -Who controls knowledge flows. -Awareness of node
Closeness centrality	High	-Closeness of a node with other nodes -Cohesion	-Direct bargaining and exchange with other actors
Betweenness centrality	High	- Number of times a node acts as a bridge along the shortest path between two other nodes. -Importance of a node	-Brokering contacts among actors to isolate them or prevent connections -Controls flow of information
Eigenvector Centrality	High	-Influence of a node in a network -Global prominence of a node	-Role of a node -Power of a node -Who controls knowledge flows. -Awareness of node
Centralization	High	-Dominance of a node -Global prominence of a node	-Role of a node -Power of a node -Who controls knowledge flows. -Awareness of node
Homophily	High	- The tendency of individuals to associate and bond with similar others - Cause of relationship Strength	-Community Formation - Influences diffusion patterns over a social network
Robustness	High	-Clustering of individuals is high for a given network -Important node & links of network	-Connectivity of network -Cluster Formation
Efficiency	High	-Can access a large number of different nodes, through a relatively small number of ties -Average Path length decreases.	- Access of resources - Reachability -Performance of the network
Effectiveness	High	-If the cluster of nodes that can be reached through non-redundant contacts are more, so that center node get more information.	- Center node Awareness and opportunities -Information flow
Diversity	High	- More non-redundant paths to reach the target node	- Performance of the network - Alternative opportunities.

Table I : Shows comparison between various social network measures

VI. Conclusion And Future Scope

The work presented in this paper seeks to inform network scientists, researchers, and practitioners regarding network measures, their dependent parameters, effect & influence of network measures on whole network. These measures have important role in deciding the network structure and network dynamics. There importance increases further, as they have significant importance in the field of social network analysis which focuses on the analysis of patterns of relationships among people, organizations, states and such social entities.

Networks structural and networks performance measures are discussed in this paper. We try to analyze the effects of ties and the networks in which they are embedded: Is it the nature of dyadic ties (e.g., strong/weak; kin/friend) or of the networks (e.g., large/small; densely/sparsely knit; clustered/integrated) that affects the kind and quantity of resources that flow through these networks. We also analyzed the local and global influence of a node with measures like degree, distance, centrality etc. The local and global measures defined are powerful generalizations of degree, closeness, betweenness and eigenvector centrality. These centrality measures have the potential to uncover important information regarding network subset properties. However, their importance vary depending on the ground truth measures and networks considered. The similarity of network members to each other (by some criteria) and to the person at the center of the network is uncovered with the measure homophily which largely affects the formation of communities not only nearby distances but with large distances too. Network performance measures are studied to know the factors which affects the networks robustness, efficiency, effectiveness and diversity and how these measures contributes in improving the performance of a network. Finally a comparison among all different measures and the results produced by these measures is done to know about similarities and dissimilarities between them and to outlined the indications, effects and influence of these measures on other nodes, groups and the whole network. Further analysis of this paper gives the answer of some important question of the social network field.

In future we will work on the mathematical part of these measures, so that these measures help the analyst to make more informed decisions. There is certainly still room for work and improvement on the speed and sensitivity of community structure algorithms. We also like to work on clustering efficiency and computational efficiency of the networks for increasing speed of community structure detection algorithms.

VII. References

[1] S. Wasserman and K. Faust, " *Social Network Analysis: Methods and Applications*," Cambridge, ENG and New York: Cambridge University Press, 1994.
[2] P. Carrington, J. Scott, and S. Wasserman , " *Models and Methods for Social Network Analysis,"* New York: Cambridge University Press, 2004.
[3] Butts, T. Carter, “Social network analysis: A methodological introduction,” *Asian Journal of Social Psychology* 11(1), 2008, pp. 13-41.
[4] J. Coleman, "Foundations of Social Theory," Harvard, 1990.
[5] Freeman and Linton, “Centrality in social networks: Conceptual clarification,” *Social Networks* 1, 1979, pp. 215-239.
[6] Borgatti, Steve and Martin Everett , “Notions of position in network analysis,” *Sociological Methodology* 22, 1992, pp. 1-36.
[7] K. Faust, "Very Local Structure in Social Networks." *Sociological Methodology*, 37, 2007, pp. 209-256.
[8] Blau, M. Peter "A Theory of Social Integration," The American Journal of Sociology, (65)6, 1960 (May), pp. 545-556 .
[9] M.E.J. Newman, "Networks: An Introduction," Oxford, UK: Oxford University Press, 2010.
[10] Borgatti, P. Stephen, "Centrality and Network Flow," *Social Networks* (Elsevier) 27, 2005, pp. 55–71.
[11] L. Freeman, "Graphic techniques for exploring social network data," New York: Cambridge University Press, 2005, pp. 248-269 .
[12] M. Granovetter, " The strength of weak ties," *American Journal of Sociology* (78), 1973, pp. 1360-1380.
[13] P. Bonacich, "Power and centrality: a family of measures," American Journal of Sociology 92, 1987, pp. 1170-1182.
[14] Izquierdo, R. Luis & Robert A. Hanneman, "Introduction to the Formal Analysis of Social Networks Using Mathematica," Available: http://www.luis.izquierdo.name. Burgos, Spain, 2006.
[15] R.A. Hanneman and M. Riddle," Introduction to Social Network Methods," Riverside, CA: University of California, Riverside, 2005 Available: http://www.faculty.ucr.edu/~hanneman /nettext/ .
[16] Blau, M. Peter,"Exchange and power in social life, "NewYork: John Wiley, 1964.
[17] A.K.Chatterjee, S.Bandyopadhyay,& A.R. Rao, "Relative importance of different factors for boundary of reciprocity,". Connections: Bulletin of the INSNA, 16, 1993, pp. 15–22.
[18] Scott, "Social Network Analysis. A Handbook", Sage Publications, 2000.
[19] T. Snijders, "The Degree Variance: An Index of Graph Heterogeneity," Social Networks 3, 1981, pp.163–174.
[20] A.L. Barabási., "Linked: The New Science of Networks," New York: Perseus, 2002.
[21] S. Wolfram ,"A New Kind of Science," Wolfram Media, 2002.
[22] M. McPherson, Smith-Lovin, & J.M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* 27, 2001 pp. 415–444.
[23] P.F. Lazarsfeld, and R.K. Ronkeylaf Merton, (1954)"Friendship as a Social Process: A Substantive and Methodological Analysis". In Freedom and Control in Modern Society, Morroe Berger, Theodore Abel, and Charles H. Page, eds. New York: Van Nostrand, 18–66.
[24] P.V. Marsden," Homogeneity in confiding relations," Soc. Networks 10, 1988, pp. 57–76
[25] N.P. Mark, "Culture and competition: Homophily and distancing explanations for cultural niches". *American Sociological Review* 68, 2003, pp. 319–345. doi:10.2307/1519727.
[26] M.D. Choudhury, ‘‘Birds of a feather’’: Does user homophily impact information diffusion in social media". *Arxiv preprint*, 2010, pp. 1–31.
[27] N.A. Christakis, J.H. Fowler, "The spread of obesity in a large social network over 32 years," (PDF). *The New England Journal of Medicine* 357, 2007, pp. 370–379. doi:10.1056/nejmsa066082.
[28] D. Centola, R. Willer, M. Macy "The emperor’s dilemma: A computational model of self-enforcing norms. American Journal of Sociology, 110, 2005, pp. 1009–1040.
[29] Mustafa, Gönenç, "Impact of Homophily on Diffusion Dynamics Over Social Networks" (PDF). *Social Science Computer Review*, May 29, 2014.
[30] R. Burt,"Structural Holes: The Social Structure of Competition," Cambridge, MA: Harvard University Press; 1992.

Himanshu Sharma recieved his B.E degree in computer science from University of Rajasthan in year 2006. He is pursuing M.Tech presently. His main research interests includes data mining, network analysis and knowledge engineering.

Vishal Srivastava recieved his B.Tech & M.Tech in degree computer science from RGPV college in 2004 and 2007 respectively and pursuing his PhD presently. He is working as professor in Arya College of Engineering and Information Technology, Jaipur, Rajasthan. He is member of CSI, IAE, UACEE, SCIEI, IJCTT, IJATER. He is also editorial board member and reviewer in IJACR, IJRET, IJCST, IJCT, IJAERD, IJATEE etc in area of CS and IT. His area of interest includes Networking and Data Mining. He has published 76 International papers and 24 National papers in international journals, conferences and national conferences.He also get grants from DST, Government of Rajasthan for his R&D and projects like SAMADHAN, Vaccination schedule alert system on mobile, Petition signing an android application.