REVIEW

Analyzing negative ties in social networks: A survey

Mankirat Kaur [*], **Sarbjeet Singh**

UIET, Panjab University, Chandigarh, UT, India

Abstract   Online social networks are a source of sharing information and maintaining personal contacts with other people through social interactions and thus forming virtual communities online. Social networks are crowded with positive and negative relations. Positive relations are formed by support, endorsement and friendship and thus, create a network of well-connected users whereas negative relations are a result of opposition, distrust and avoidance creating disconnected networks. Due to increase in illegal activities such as masquerading, conspiring and creating fake profiles on online social networks, exploring and analyzing these negative activities becomes the need of hour. Usually negative ties are treated in same way as positive ties in many theories such as balance theory and blockmodeling analysis. But the standard concepts of social network analysis do not yield same results in respect of each tie. This paper presents a survey on analyzing negative ties in social networks through various types of network analysis techniques that are used for examining ties such as status, centrality and power measures. Due to the difference in characteristics of flow in positive and negative tie networks some of these measures are not applicable on negative ties. This paper also discusses new methods that have been developed specifically for analyzing negative ties such as negative degree, and h^* measure along with the measures based on mixture of positive and negative ties. The different types of social network analysis approaches have been reviewed and compared to determine the best approach that can appropriately identify the negative ties in online networks. It has been analyzed that only few measures such as Degree and PN centrality are applicable for identifying outsiders in network. For applicability in online networks, the performance of PN measure needs to be verified and further, new measures should be developed based upon negative clique concept.

© 2015 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author. Tel.: +91 9463543880.
E-mail addresses: mannsunshine09@gmail.com (M. Kaur), sarbjeet@pu.ac.in (S. Singh).

Contents

1. Introduction

Online social networks are becoming popular among large number of people, as a source of forming virtual communities online. These communities are developed by creating profiles and maintaining personal contacts of each user through social interactions. Schneider et al. [1] defined social network as "OSNs form online communities among people with common interests, activities, backgrounds, and/or friendships. Most OSNs are Web-based and allow users to upload profiles (text, images, and videos) and interact with others in numerous ways". Social interactions in these networks are crammed full with positive and negative relations. Positive relations are formed by support, endorsement and friendship and thus, create a network of well-connected users which is useful for the promotion of market products, brands, services and new research ideas on social media [2]. Negative relations, on the other hand, are a result of opposition, distrust, antagonism and avoidance. Negative relationships represent a persistent, recurring set of negative social intentions toward another person. Negative relations are central to clusterability theory [3], balance theory [4,5], their generalizations [6], semigroup work [7] and social network analysis [8]. Many areas of negative networks need extensive research such as bullying, group conflicts and social exclusion. Because of the increase in criminal activities in online networks such as people masquerade others

by creating fake profiles and conspire against business rivals by presenting fake reviews of products being promoted on these networks, there is a need to identify these negative actors of network by observing their connection patterns. In another example, one political party opposes the policies and efforts made by other parties on these networks by commenting against them and exaggerating their own policies by pretending them to be in welfare of public. Therefore, these undesirable activities of political parties need to be monitored.

Many researchers have focused on studying negative ties in different fields of research. Bohn et al. [9] have analyzed the access to social capital in OSNs by examining the actual communication ties (positive and negative) among actors, whereas, Box-Steffensmeier and Christenson [10] studied the interest group coalition by considering positive and negative ties. Smith et al. [11] have accessed the phenomenon of homophily in social networks by studying network ties of both positive and negative nature. De Jong et al. [12] analyzed the impact of negative relationships on attitudes of each team member and performance in teams. Recently, many researchers have dedicated their research on investigating negative ties to explore different concepts in social networks such as studying attitude diffusion [13] and analyzing aspects of inter-ethnic relationships between secondary school students [14].

The interest of this paper is in the relations that are themselves negative rather than negative consequences of positive relations. There are many standard datasets depicting such kind of relations such as, skirmish relation in "bank wiring room" data reported by Roethlisberger and Dickson in 1939 [15]; enmity "hina" relation reported by Read in 1954 [16] of social network of tribes of the Gahuku–Gama alliance structure of the Eastern Central Highlands, New Guinea; and the disesteem and disliking relations among group of monks reported by Sampson in 1969 [17]. The same types of relations are also present in today's online networks such as Wikipedia, Epinions, and Slashdot which are online rating websites. In Wikipedia, users can vote for or against other users in elections to nominate them for admin status. If one user gives supporting or opposing vote to another user, then it leads to creation of positive or negative link respectively [18]. Epinions.com is a common consumer review website which forms an online social trust network where users create signed relations of trust or distrust with each other. Members of site can give positive or negative ratings to products of website as well as rate the reviews given by other members [19,20]. The visitors of site can check new and old reviews of product and then decide which product to purchase. The links between nodes and persons who give reviews about products are explicitly labeled as positive or negative. Slashdot is a technology related news website that entails specific user community. The news related to current technology is submitted by users, in which editors of site evaluate and provide open discussion among readers. Each news article also has section for comments attached to it, where users of site post their comments and lead to open threaded discussion. The comments associated with each news article are then rated by editors of site by using moderation system. The moderators rate each comment as $+1$ or -1 [21–23]. In 2008, Slashdot launched a new feature known as Slashdot Zoo which allows users to tag other users as friends (positive link) or foes (negative link). All these represent negative ties between actors in a group. It has been found that negative ties are less in number and are generally treated in the same way as the positive ties in most of the concepts such as blockmodeling analysis [24–26].

Now, question arises that whether the standard methods used for analyzing positive ties in networks can also be used for studying negative ties or not? This involves systematic analysis of standard techniques and changes in interpretation of these techniques to make them applicable for negative ties. There ascends many concerns when dealing with negative tie data:

- *Transitivity:* In positive relations, ideas or information is quickly delivered to group of nodes, directly or indirectly connected to a node usually covering distance of two or more length path depicting high level of transitivity, whereas in negative relation, flow does not follow this pattern. There is hardly a diffusion of information to more than one length path showing low level of transitivity.
- *Sparseness:* Positive ties form very dense networks but negative networks are very sparse and usually form highly disconnected graphs which make centrality approaches (to find most central and influential node in network through which most of the traffic flows) difficult to apply.

For example, in positive network, information can easily diffuse from m to p through directed path $m \rightarrow n \rightarrow p$ where m is connected to n which is further connected to p but this is not possible in negative networks. Because of these few reasons, not all the standard approaches are applicable to negative relations.

Therefore, in this paper many new recent techniques that are used for examining only negative ties are discussed. For example, by using any strong negative node centrality measure, we can analyze whether a person is chosen or ignored for promotion in trust network of organization based upon high or low score calculated through this measure [27]. Similarly, potential victims of bullying can also be identified through scores calculated by these measures and can plan interpolation in early stages. In practical situations, both the positive and negative ties exist in social networks. So, in order to analyze the effect of these ties jointly on centrality scores of actors, we must have measures that can examine both types of ties simultaneously. Therefore, various such types of techniques are explored, analyzed and compared on the basis of various factors such as weightage of direct and indirect links, concept used for analyzing ties and complexity. The focus of this paper is to survey the social network analysis approaches for exploring the techniques that are suitable for identifying the outsiders or negative nodes of network. The study of identifying negative ties till date is carried out on small datasets of offline networks. Analyzing these ties on datasets of online social networks is still an unexplored area.

This paper is organized as follows. Section 2 discusses the standard methods of analyzing ties and checks their applicability on negative ties; Section 3 consists of studying negative ties through influence measures. Some new projected approaches for interpreting negative ties by using standard measures are discussed in Section 4. New measures of negative ties are analyzed in Section 5 followed by discussion and comparisons of various mixed data measures in Section 6. The strengths and weaknesses of all the discussed measures are summarized in Section 7. Finally, the paper is concluded with some new future directions in current field. An attempt has been made to cover important studies made in this field but study in this paper shall not be considered exhaustive in any sense.

2. Standard methods

2.1. Analyzing ties through equivalence concepts

Network is defined as a set of ordered pair of nodes exhibiting relations. It is depicted by a graph, more generally a digraph (directed graph). Relation is a specific type of tie existing between nodes of graph usually represented by aRb and path of tie is from a to b. To trap the intermingling of role relations in social process or structure, the social ties are examined as follows [28]:

a. Generation and elimination of social ties are commenced by other ties existing in network.
b. The impulses generated by individual nodes in network lead to activation of social process (communication among population).

c. The nature of tie between pair of persons or nodes depends upon pattern that is followed by ties of that type in network.

Symmetry is an important property of social relations which says that for any nodes m, n and relation R, if mRn then there must exist nRm. Friendly relations in social network are also symmetric that is "Receiving and acknowledging the friendly expectations projected to one by friend, is treated as an intrinsic part of one's own projection of same friendly expectations to him, so long as all pairs in the population are symmetric in this way" [28]. Symmetry is also present in negative relations. Equivalent nodes in social network can be described as nodes that are having similar pattern of relations with other nodes of network and thus playing same roles and positions. There are primarily three types of role or positional equivalence exists in network as briefed below.

2.1.1. Structural equivalence

Different authors have defined structural equivalence of nodes in different ways. Lorrain and White [28] gave categorical approach in which relations between the nodes that can be compounded are called morphisms. Different types of relation between nodes are coupled to form compound relation, for example, one's sister's friend or one's batch mate's friend, and these are composite of two relations R and S where R is sister in first case and batch mate in second case and S is friend in both cases. The compound relation for any nodes a, b, c is defined as $a(RS)c$, if and only if aRb and bSc exist, where, R and S relations are morphisms. Composition operations on R and S to form RS and its graph on the set of nodes together are known as a category. Category C constitutes 2 types of classes CObj and CMor. Elements of class CObj are the nodes in network also called objects and elements of CMor are the morphisms exist between pair of nodes. They defined concept of category as composed of three stages: the objects, the morphisms and composition of morphisms. On the basis of this concept, Lorrain and White [28] defined Structural Equivalence as "Objects a, b of category C are structurally equivalent if, for any morphism M and any object x of C, aMx if and only if bMx, and xMa if and only if xMb." This means any nodes a, b are structurally equivalent if node a relates to any other node x in the network with relation M which also exists between b and x. Thus, both the nodes a, b are completely equivalent and thus are substitutable [28].

Another definition of structural equivalence was given by White and Reitz [25] in which they coined a concept of homomorphism as the mapping of points or nodes of graph to the actors or nodes in the image of the graph. They defined structural homomorphism as "There are some types of roles in networks in which it is expected that all the occupants of that role should be identically connected to occupants of counterpart's role" [25] i.e. nodes a and b must belong to role 1 if they are having same image after mapping through function f:

$$\left.\begin{array}{c} a \\ b \end{array}\right\} \xrightarrow{f} x = f(a) = f(b)$$

As well as d, e belong to role 2 as:

$$\left.\begin{array}{c} d \\ e \end{array}\right\} \xrightarrow{f} y = f(d) = f(e)$$

Suppose, if a node a is related to a node, say d and b are related to node e, through a relation R then:

a	R	d
b	R	e

(belongs to one role)   relation   (counterpart's role)

Also d, e are counterpart roles of a, b if aRd, aRe, bRe, and bRd exist in network. Now structural homomorphism states that all the members of role a, b should be related to all the members of counterpart's role d, e as well as to their own members [25]. This can be represented pictorially as follows (see Fig. 1).

The equivalence brought by structural homomorphism is structural equivalence. Structural equivalence is then defined as "If $G = \langle P, R \rangle$ where P is the finite set of points or nodes of graph G and R is the relation on set of ordered pairs of points or nodes, and $\equiv$ is an equivalence relation on P then $\equiv$ is a structural relation if and only if for a, b, c ε P where $a \neq b \neq c$, $a \equiv b$ implies [25]:

i. aRb if and only if bRa;
ii. aRc if and only if bRc;
iii. cRa if and only if cRb and
iv. aRa implies aRb."

According to this definition, structurally equivalent nodes are related to each other with same relation R, which also relates them to other nodes of network.

Yet, third definition is given by Everett and Borgatti [26] as two vertices or nodes a and b of digraph are considered structurally equivalent if and only if $N_i(a) = N_i(b)$ and $N_o(a) = N_o(b)$. Here, $N_i(a)$ is the in-neighborhood of vertex a, i.e. a set of vertices or nodes from which a receives connections; $N_o(a)$ is the out-neighborhood of vertex a i.e. the set of vertices which receives connections from a. They described that if in and out-neighborhood set of any two vertices is same then those vertices are structurally equivalent. All the three descriptions explained same concept which can be briefed as: "Any nodes m, n are structurally equivalent if both nodes are having same type of ties to same nodes in network."

Clearly, this equivalence of nodes can possibly occur in positive ties network such as school environment in which two teachers say, A and B teach same set of students $\{E, C, D\}$ in same school, then A and B are structural equivalent nodes in network of school with positive tie existing between teacher and students. Here, if we interchange their positions as they are teachers in same network of students there is no effect on structure of network. In case of negative ties such as bullying, it is possible to have any nodes A and B bullying a
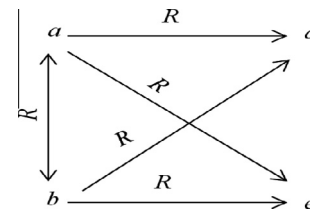
Figure 1   Structural homomorphism, a b belongs to one role and d and e are counterpart's role.

group of students $\{M, N, P\}$ who are potential victims. Now, A and B are considered to be structurally equivalent as they are having same ties with same group of students and $\{M, N, P\}$ are structurally equivalent to one another as victims of bullying having abhorrent relation with candidates of bullying $\{A, B\}$. These nodes A and B are expected to have similar roles and positions in the network. Thus this equivalence is also possible in negative ties network.

2.1.2. Regular equivalence

This concept is used in graph theory for nodes which are showing equivalence in terms of their connections and relations to other nodes or actors in a graph. Many authors have defined regular equivalence by considering different concepts used in graph. This term was first used by White and Reitz [25] using concept of homomorphism. They stated regular homomorphism as "In role systems it is expected that all the occupants of one role should be identically connected to occupants of their counterpart's role" [25]. Here, each member should be connected to at least one member of counterpart's role but condition of connection to every member of counterpart role, as in case of structural homomorphism, is not necessary in this type. In continuation with previous example, regular homomorphism only requires:

$$a \xrightarrow{R} d$$
$$b \xrightarrow{R} e$$

Cross connection of a to e and b to d in this mapping is not necessary. The equivalence brought by regular homomorphism is regular equivalence which was defined as "If $G = \langle P, R \rangle$ where P is a set of points and R is relation between ordered pair of nodes, $\equiv$ is an equivalence relation on P, then $\equiv$ is a regular equivalence if and only if for all $a, b, c \varepsilon P, a \equiv b$ implies [25]:

    i. aRc implies there exists $d \varepsilon P$ such that bRd and $d \equiv c$ and

    ii. cRa implies there exists $d \varepsilon P$ such that dRb and $d \equiv c$."

According to this, points in graph are considered regular equivalent points if they are related to their corresponding equivalents with same relation R. Everett and Borgatti [26] used concept of regular coloration to highlight various results known about regular equivalence. They defined directed graph or digraph as $d(v, e)$ where v is set of vertices and e is set of edges and also introduced in-neighborhood, out-neighborhood of vertex j as $N_i(j)$ and $N_o(j)$. Coloration in digraph D was described as allocating different colors to vertices. Regular coloration exists if and only if for all $m, n \varepsilon v$: $C(m) = C(n)$ implies [26]

    i. $C(N_i(m)) = C(N_i(n))$ and
    ii. $C(N_o(m)) = C(N_o(n))$

That is, if the color of vertices in digraph is same then its in-neighborhood and out-neighborhood must have same set of colors. Such coloration is regular in nature and brings regular equivalence among nodes of graph. They also used image digraph d' to generate set of equivalent classes which contain regular equivalent vertices. The image digraph is $d'(C(v), e')$ where $C(v)$ is a set of colors of vertices in d as labels of vertices

in d' and e' are set of edges that exist between pair of vertices if and only if nodes of that colors in d are adjacent to each other [26]. For example, consider a graph of five vertices connected to each other through directed links in Fig. 2. Vertices with overlapping set of in- or out-neighborhood vertices are colored with same color and rest with different colors. The name of vertex is written inside circle and color of vertex is written outside. Above coloration is regular as the colors of in-neighbors of 4 and 2 are red and yellow and those of out-neighbors are yellow for both. Similarly, colors of in-neighbors of 1 and 5 are red and yellow and colors of out-neighbors are green and red for both. Image digraph D is shown in Fig. 3.

There is an edge from vertex red to green because node 1 is connected to node 2 and node 5–4. This image digraph divides 5 vertices into set of 3 equivalent classes in which every node is regularly equivalent to each other (for example node 1 and node 5).

Both the statements explained that regular equivalent points or vertices in graph are the one that are connected to their corresponding equivalents with single relation R. In first case, when R is a positive tie, like father and son relationship, then possibly there can exist regularly equivalent nodes or persons a and b belonging to regular set of fathers such that a is a father of c and b is father of d, where c and d are the members of regular equivalent set of sons. In case of negative tie network where nodes are negatively connected, like gossip network at workplace in which A gossips about C and B gossips about D due to some conflicts over a project directed
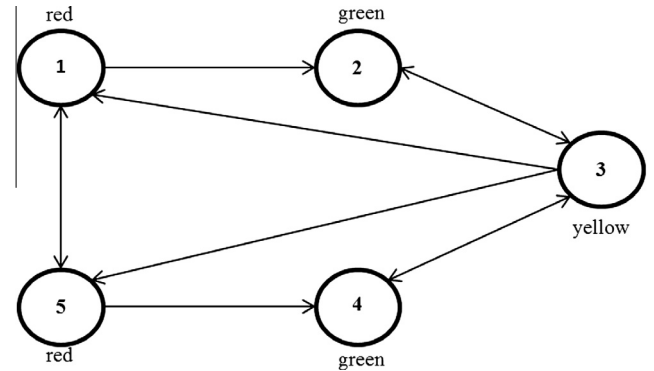
Figure 2 Colored graph with name of vertices written in circles and color of vertex written outside circle. Regular equivalent nodes are colored with same color.
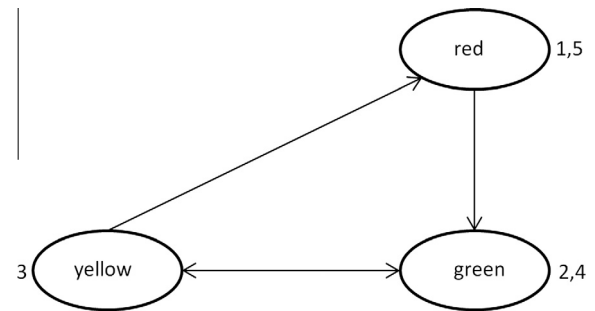
Figure 3 Image digraph with name of vertices written inside ovals is colors of vertices of digraph written outside ovals.

by all of them. Here A and B are both acting as regular equivalent nodes having similar type of relations to their equivalents C and D in a group.

2.1.3. Automorphic equivalence

The concept of automorphism was first used by Lorrain and White [28] in context of social networks. They used the term endomorphism, which are self-loops in directed graph $d(v, e)$ representing the node's consciousness about its locus in social network structure. Endomorphisms were used as an important notion to identify group of nodes with which a node in network is most likely to be associated [28]. The more desired statement was given by Everett and Borgatti [26], "Automorphism is any permutation Π on set of vertices v so that adjacency will remain preserved after permutation", i.e. $(m, n) \, \varepsilon \, e$ (set of edges in d) if and only if $(\Pi(m), \Pi(n)) \, \varepsilon \, e$. Consider the graph in Fig. 2, $\Pi(1) = 5$, $\Pi(2) = 4$, $\Pi(4) = 2$, $\Pi(5) = 1$. Then automorphic equivalence can be stated as "two vertices m and n of digraph D are considered automorphically equivalent if and only if there exists an automorphism Π such that $\Pi(m) = n$" [26].

This notion of equivalence can be used in positive network, consider two branches of school as social network with A as a principal of group of teachers $\{M, N, P\}$ in first branch and B as a principal of group of teachers $\{U, V, W\}$ in second branch. D is the director of both branches of school. A and B are automorphically equivalent to each other as shown in Fig. 4. If director decides to move A to branch 1 and B to branch 2, then this transfer cannot take place without exchanging group of teachers in order to preserve relations in structure of network. Similarly, in negative ties network of bullying, it is possible that nodes A and B bullying two different groups of pupils are considered automorphically equivalent and if we permute their positions in network along with group of pupils the structure of network remains unchanged.

From this discussion it is clear, that all general concepts and theories of equivalence of nodes are applicable to both positive and negative ties and help in analyzing the pattern of flows and structure of nodes in network.

2.2. Analyzing ties through statistical techniques

As all equivalence concepts are equally applicable to both types of ties, there exist other methods known as statistical
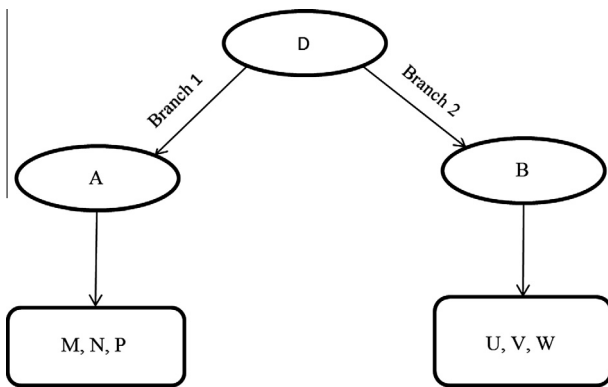
**Figure 4** Network of School branches, in which nodes represent the designation at different levels.

techniques such as QAP correlation [29], regression [30], and ERGM. that are used to check the similarity or correlation among various measures. These approaches could be used for analyzing both types of ties but the model required for correlation may differ. For example, positive relation of friendship with colleagues within organization can be positively correlated with sameness of states to which they belong as well as with promotion on job as these colleagues may help in providing critical information and references. Similarly, negative relations of disliking with colleagues at job can be positively correlated with demotion in organization structure as colleagues may provide negative feedback about a person to higher authority. However, the sameness of state to which they belong would not necessarily be positively correlated with negative relations. The only difference between the required models of correlation of both ties is promotion and demotion of an employee. Rest all factors such as sameness of language, state or region, and age can be modeled in same manner for both negative and positive relations. The exponential random graph model (ERGM) framework is also suitable for negative ties but the sparseness of negative data matrix causes problem in calculation of correlation and makes ERGM software packages less relevant [31]. Therefore, different models or configurations should be developed for better analyzing negative tie networks.

2.3. Analyzing ties through centrality measures

The word 'centrality' refers to the central element, item or actor in any domain of knowledge. The usage of term centrality in different domains by researchers refers to finding important elements which reflect significant properties of domain. The centrality was used in human communication paths firstly by Bavelas [32]. The research experiment of centrality was conducted at Group network Laboratory, M.I.T. in 1940s by Leavitt [33] and Smith [34]. The results showed that efficiency of group in problem solving is affected by centrality [35]. Despite of its application in group problem solving, it was also used by Cohn and Marriott [36] to integrate large, heterogeneous and diverse social culture of country such as India and to administer its social cultures politically. After this, the concept of centrality was analyzed by Pitts [37] for its usage in urban development by reconstructing river transportation network in central Russia. Centrality can also be used in design of large organization by combining central units of various small organizations to form one large unit as examined by Beauchamp [38] and Mackenzie [39]. Further Rogers [40] suggested that centrality for organization could be computed from the characteristics of the firm and properties of communication network of firm. Beside these application areas, concept of centrality plays an important role in social networks. It was used to find most central and influential node in network through which most of the traffic flows and this node controls what type of information should diffuse into network. The measures to find centrality of node depend upon structural properties of network and they make use of flows to examine these characteristics [41].

Various terms are used in centrality measures that can be briefed as follows:

Consider Fig. 5: it depicts a graph consisting of labeled points and edges. This graph can be visualized as a social
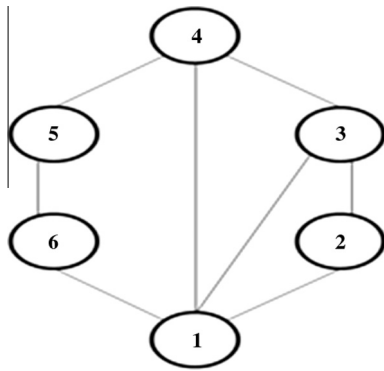
Figure 5 Graph of a network with vertices depicting nodes and edge referring to communication link between the pair of nodes.

Figure 6 Star network of nodes. Node 1 has highest value of centrality possible in any network of nine nodes.

network in which persons are connected to each other through edges of graph.

i. *Node:* a point in a graph corresponds to each person or actor in network. Like 1, 2, 3 etc.
ii. *Degree:* maximum number of connections from a given point to all other points in network. For example, degree of point (3) = 3 and point (1) = 4.
iii. *Link:* an edge in a graph corresponds to each communication link that connects a pair of persons in network.
iv. *Path:* sequence of edges between pair of points. For example: path between 5 and 2 is 5 → 4 → 3 → 2, and these three edges constitute a path.
v. *Cycle:* a path starting from and ending at same point like 1 → 2 → 3 → 1.
vi. *Connected graph:* When each point is reachable from every other point in graph.
vii. *Geodesic:* a shortest path between a pair of points, like four paths between 1 and 4 exist: 1 → 6 → 5 → 4; 1 → 2 → 3 → 4; 1 → 3 → 4; 1 → 4. From these paths last one 1 → 4 is geodesic.

There are broadly two types of centralities categorized by Freeman [41]:

a. *Point centrality:* Point centrality can be described by central position of star or hub which is considered as the most central position possible in any network, like position of node 1 in Fig. 6. According to Freeman "A person located in the center of a star is universally assumed to be structurally more central than any other person in any other position in any other network of similar size" [41]. And thus it came out that central point of star should possess three properties to become most central, that are [41]:
   i. It should have maximum degree in the network.
   ii. It should lie on maximum possible number of geodesics between different pair of points in network.
   iii. It should be maximally close to all other points by locating at minimum distance from them.

b. *Graph centrality:* Graph is basically a set of points depicting a social network. The centrality concept is extended to be used in graph and is associated with compactness of graph. Freeman defined compactness as "A
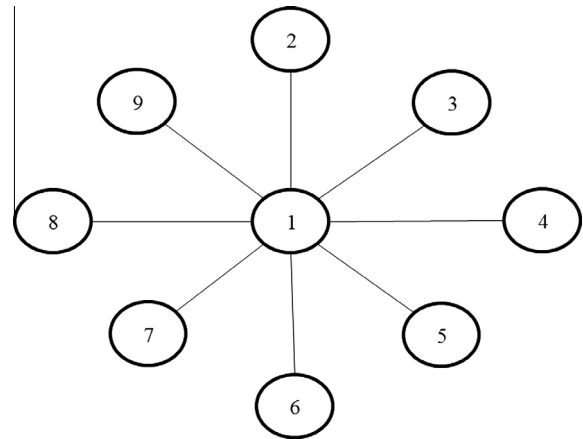
graph should be compact to a degree that the distances between pair of points are short and this graph theoretic conception is extended to social networks and renamed as graph centrality" [41]. Graph centrality measures are based upon difference between centrality of other points and the most central point in the network or graph. Freeman [41] suggested that measures of graph centrality must contain some features:
   i. They must include the value by which centrality of most central node exceeds the centrality of other nodes.
   ii. They must express the ratio of computed excess to the maximum possible value of differences in centrality of nodes.

Expression of measures of graph centrality must satisfy the following condition:

$$C_X = \frac{\sum_{i=1}^{n} C_X(p^*) - C_X(p_i)}{max \sum_{i=1}^{n} C_X(p^*) - C_X(p_i)} \quad (1)$$

$C_X(p_i)$ = centrality of node i. $C_X(p^*)$ = largest value of centrality of any node in network.

Denominator gives the maximum value of sum of difference between centralities for a graph of n nodes. Both of point and graph centralities can be computed by using three measures: Degree, Betweenness and Closeness.

2.3.1. Degree centrality

Point centrality was first defined in terms of degree of a point as "count of number of points to which a given point is connected or adjacent to". Referring to Fig. 5, degree of node 3 is three, degree of node 5 is two and so on. This conception of degree was introduced by many researchers as mentioned in Table 1. They all defined degree as point centrality in terms of human social networks as "person who permits direct connection to most other persons in social networks should be seen as focal point of communication where information of network flow through" [37,40,42–49]. But this definition was further realized useless as it did not consider graph size or longest geodesic in the graph that can be used to compare point centrality score over different graph sizes. Nieminen

Table 1 Comparisons of different measures of centrality.

Type of centrality	Proposed by	Year	Field	Description of measure	Usage	Type of flow	Type of walk structure
Degree centrality	M.F. Shaw	1954	Analyzing group structure	It defines centrality as the number of direct links of 1-length that a node has with their neighboring node	Concerned with communication activity in network	Parallel Duplication. For Example: E-mail Broadcast, Attitude influencing, Money Exchange	Path, Trail, Walk
	C. Faucheux and S. Moscovici	1960	Social influence				
	W.L. Garrison	1960	Interstate highway				
	K.D. Mackenzie	1966	Communication networks				
	F.R. Pitts	1965	Communication paths in urban development				
	D.L. Rogers	1974	Inter-organizational relations				
	J.A. Czepiel	1974	Diffusion of technological innovation				
	J. Nieminen	1973 1974	Centrality in graph				
	Y. Kajitani and T. Maruyama	1976	Assessment of communication networks				
	L.C. Freeman	1979	Centrality in social networks				
Betweenness centrality	A. Bavelas	1948	Group structure	It defines centrality as number of shortest paths between pair of nodes on which a given node lie	Concerned with control of communication	Transfer e.g. Packet Delivery system	Geodesic
	M.E. Shaw	1954	Behavior of individual in small group				
	B.S. Cohn and M. Marroit	1958	Integration of Indian Civilization				
	L.C. Freeman	1979	Centrality in social networks				
Closeness centrality	A. Bavelas	1950	Communication patterns in groups	It defines centrality of a node as the inverse of sum of minimum shortest paths from all other nodes in network	Concerned with independence and efficiency in spreading message	Transfer e.g. Packet Delivery	Geodesic
	M.A. Beauchamp	1965	Efficiency of organization				
	G. Sabidussi	1966	Centrality index in graph				
	R. L. Moxley and N. F. Moxley	1974	Uncontrived social network			Parallel Duplication e.g. gossip process	Path Trail Walk
	D.L. Rogers	1974	Inter-organizational relations				
	L.C. Freeman	1979	Centrality in social networks				
Eigenvector centrality	P. Bonacich	1972	Social networks	It defines centrality of node as proportional to the sum of centrality of neighboring nodes	Concerned with Influencing System	Parallel duplication e.g. Attitude Influencing	Unrestricted walks

Four measures of centrality are compared on the basis of their usage in performing specific activities, type of flow they follow and type of walk structure permitted by each measure to flow things through network.

[48] provided definite measure of degree centrality in terms of number of adjacencies as

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k) \tag{2}$$

where $a(p_i, p_k) = 1$ if p_i and p_k are connected else 0. This measure is partially a function of size of network. But the absolute count of activity of node may not be desirable for most of applications. So, in order to remove effect of network size, it was normalized by Freeman [41] to get relative centrality i.e.

$$C_D' = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1} \tag{3}$$

Degree centrality indicates the potential communication activity of a node in network. Graph centrality in terms of degree of graph was given by Freeman [41] as

$$C_D = \frac{\sum_{i=1}^{n} [C_D(p^*) - C_D(p_i)]}{n^2 - 3n + 2} \tag{4}$$

The maximum possible value of denominator is $(n-2)(n-1) = n^2 - 3n + 2$ where n is the number of nodes in network. Degree centrality is the measure to calculate immediate risk of getting infected in network of infected nodes in one time period from directly connected infected node [50]. As shown in Table 1, degree measure is used in parallel duplication flows and thus appropriate for walk based processes [51]. In positive tie network, degree measures are used to calculate popularity of node i.e. the node that is receiving more number of positive ties from other nodes in network (high in-degree) and sending many ties to other nodes (high out-degree) is the most liked person in network. Similarly, interpretation can also be made in negative ties network of disliking relationships to identify the most disliked individual in the network who is receiving more number of negative ties from all other individuals. Degree based graph centralization can be used in negative tie network to find the decrease in group cohesion due to the presence of negative relations of disliking and distrust. Degree centrality is the only measure out of three that can be used in both positive and negative networks because degree of a node is independent of flows and calculated on the basis of number of direct connections. However, this measure cannot calculate the centrality possessed by a node due to the presence of other nodes at more than one length path [31].

2.3.2. Betweenness centrality

Betweenness as a measure of point centrality can be stated as the number of counts with which a point lies in between the shortest paths or geodesics linking pairs of points. This measure is based upon assumption that information flows only along shortest paths connecting pair of points in network [41]. Such a node falling on most of the geodesics in network is considered to be central and controls transmission of information in group thus acts as its coordinator [36,52] For example in Fig. 6, in star network center node 1 is the most central node as it lies on all the geodesics connecting all pairs of nodes in network. Betweenness of a point can be determined easily but when there exist multiple geodesics between pair of points then partial betweenness is used in the form of probabilities [41]. For example there are two geodesics between nodes 2 and 5 in Fig. 5: $5 \rightarrow 4 \rightarrow 3 \rightarrow 2$; $5 \rightarrow 6 \rightarrow 1 \rightarrow 2$. Here, probability of using any path out of two is ø. Let's take random node say, 3. The probability of this point to occur on randomly selected geodesic path between 2 and 5 is ø * 1.

This can be mathematically written as

$$a_{ij}(p_k) = \frac{m_{ij}(p_k)}{m_{ij}} \tag{5}$$

where m_{ij} is number of geodesics between points i and j and $m_{ij}(p_k)$ is the number of geodesics between i and j that contain point p_k. Now, betweenness centrality of point p_k can be written as [41]

$$C_B(p_k) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(p_k) \tag{6}$$

where $i < j$, n is the number of nodes in network. $C_B(p_k) = 1$ when there is only one geodesic between pair of nodes that contain node p_k. As the frequency of occurrence of p_k on geodesic increases, the value of $C_B(p_k)$ increases. $C_B(p_k)$ is the absolute count which seems least interesting for most of the applications. Therefore, for getting the relative count it was normalized by Freeman as [41]

$$C_B' = \frac{2\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(p_k)}{n^2 - 3n + 2} \tag{7}$$

where $(n^2 - 3n + 2)/2$ is maximum possible value of betweenness of node p_k in a graph of n nodes. In case of graph centrality, betweenness was defined by Freeman as "average difference between centrality of the most central point $C_B(p^*)$ and all other points" [41]:

$$C_B = \frac{\sum_{i=1}^{n} [C_B(p^*) - C_B(p_i)]}{(n^2 - 3n + 2)(n-1)} \tag{8}$$

Betweenness is considered as a measure for control of communication with central point controlling the stream of information passing through network. C_B measures of betweenness were imposed to two restrictions:

i. These measures were defined on simple binary graphs which did not include quantitative attributes and generally evaluate relations in social networks such as strength of relationship based on number of interactions [53].
ii. These measures primarily considered only geodesic paths for information flow while in practical situations message is usually passed on random path or intentionally selected path by end nodes [54].

Due to above stated reasons, Freeman et al. [53] suggested a new measure of betweenness based on Ford and Fulkerson model on valued graph in which value is assigned to each edge in the graph of network. They related this value to the strength of social linkage i.e. amount of interactions specifying the time they spent with each other or any other social settings. These attributes were numbered by values or labels of edges. The concept of capacity of channel depicted by proximity function C was stated as the total amount of information that can be passed through communication links. f_{ij} is the amount of information passing through that channel link and c_{ij} is capacity of channel linking i and j. The relation between them defined by Freeman et al. is as follows [53]:

$$f_{ij} < c_{ij} \tag{9}$$

This equation infers that f_{ij} should be less than capacities of channels linking the points through direct or indirect paths.

In Fig. 7, the path from X_i to X_j is a direct path and the paths via X_k and X_m are indirect paths to reach X_j from X_i. Freeman et al. used Ford and Fulkerson model [55–57], in which originator of information is called source X_i and receiver of information is called sink X_j. This model posed two restrictions on flow between X_i and X_j:

  i. Flow coming out of X_i should be equal to flow going into X_j.
  ii. Flow coming in each intermediary node X_k or X_m connecting X_i to X_j should be equal to flow out of them.

On the basis of these restrictions $i - j$ cut sets were defined. Each cut set E_{ij} contains an edge from each and every direct and indirect path between X_i and X_j, i.e. if we remove any edge of E_{ij} from graph then X_i and X_j become unreachable. The capacity of $i - j$ cut set is total sum of capacities of each edge in the set [53]. The capacity of edge which is lowest in $i - j$ cut set is termed as minimum cut capacity and maximum amount of flow from X_i to X_j should be less than this minimum cut capacity [57]. In context of betweenness model in valued graph, X_k is standing between pair of points X_i and X_j, as seen in Fig. 7, such that maximum flow between them passes through X_k, depicted by $m_{ij}(X_k)$. $C_F(X_k)$ is the maximum flow between all the pair of points in graph that passes through X_k as defined by Freeman et al. [53]:

$$C_F(X_k) = \sum_i^n \sum_j^n m_{ij}(X_k) \quad \text{where } i < j. \tag{10}$$

To normalize this value, it is divided by total flows between all the pairs in network excluding X_k. Then above equation becomes

$$C_F' = \frac{\sum_i^n \sum_j^n m_{ij}(X_k)}{\sum_i^n \sum_j^n m_{ij}} \tag{11}$$

Through this measure centralization of graph determined by [53] is given as

$$C_F = \frac{\sum_{i=1}^n \left[C_F'(X^*) - C_F'(X_k) \right]}{n - 1} \tag{12}$$

where C_F is the average of sum of differences between centralities of most central point and other points. The measures of betweenness have focus on frequency of arrival of packets. Betweenness measure of centrality is used in different types of network flows in which it takes two assumptions. First, it
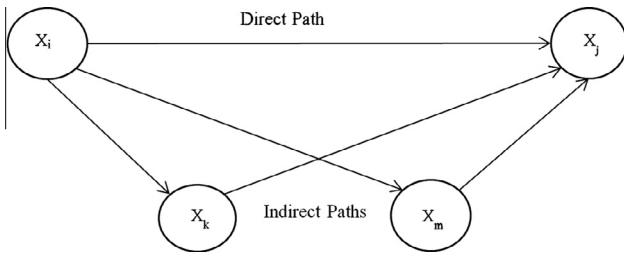
considers traffic as indivisible that can transfer from one node to another but cannot present at two places at one time through duplication. Second, traffic can take only shortest paths and flows have predefined origin and target. From Table 1, betweenness is used in packet delivery process in which packet is indivisible, present at one place at any time and can move through transfer on shortest possible path predetermined by process [51]. Betweenness measures can be used in positive ties where high level of transitivity and flow of information exist. But as discussed, flow of resources is absent in negative tie networks; therefore, it becomes very difficult to calculate betweenness measures in such type of networks [31].

2.3.3. Closeness centrality

In order to find central point or node in network, one another view can also be considered in which the point having minimum distance i.e. close to all other points in network is called central node. This node does not depend upon other nodes to deliver its message or information. Take example of star network of Fig. 6. In this, node 1 is able to deliver its message to all other nodes by itself whereas node 2 depends upon node 1 to relay its message and it is same for all other nodes [35]. This property of independence of central point is known as 'closeness' of a point. Closeness based measures were introduced by many researchers [35,38,40,58,59] as mentioned in Table 1 but the most predominant and revealing measure was given by Sabidussi [58]. He defined centrality of node as inverse of the sum of geodesics from a point to all other points in network. The measure given by Sabidussi can be written as [58]

$$C_c(p_k)^{-1} = \sum_{i=1}^n d(p_i, p_k) \tag{13}$$

$d(p_i, p_k)$ is number of edges in the geodesic from p_i to p_k and $C_c(p_k)^{-1}$ increases when distance between p_k and other nodes in network increases or closeness of p_k decreases with increasing distance and vice versa. If graph becomes disconnected then $d(p_i, p_k)$ tends to be infinite and closeness approaches zero. The $C_c(p_k)^{-1}$ is dependent on graph size and gives absolute count. So to normalize it, Freeman divided it by $n - 1$ [41]:

$$C_c(p_k) = \frac{n - 1}{\sum_{i=1}^n d(p_i, p_k)} \tag{14}$$

In terms of graph centrality closeness was measured by Freeman as [41]

$$C_c = \frac{\sum_{i=1}^n [C_c(p^*) - C_c(p_k)]}{(n^2 - 3n + 2)(n - 1)} \tag{15}$$

In case of closeness centrality, maximum closeness can be achieved in star or wheel network and minimum in complete graph in which all the points are homogenous with no difference in their centrality. Closeness can also be interpreted as expected arrival time of something flowing through network only if it flows along shortest path [51]. If raw closeness score of node is low then node is having short distance from other nodes and receives flow early. Closeness as an index of centrality is used only in processes that follow shortest paths in network, such as packet delivery process, or follow parallel duplication where things flow parallel through all possible paths, as evident from Table 1. In latter case, closeness does not indicate the reception speed as all possible paths are

Figure 7 X_i is source, X_j is sink and information between them may flow through direct path or through X_k or X_m titled as indirect paths.

traversed, including shortest paths, like in gossip process in which it cannot be determined who receives information and in what order [51]. Also closeness can be computed only in connected graph with flow-based network [41]. Therefore, this measure can easily be applied on positive ties where transitivity of flow is maximum, like friends of friends are friends. But in case of negative ties networks, which are usually represented by disconnected graphs, closeness becomes zero as nodes are not reachable from one another. And also things do not flow for more than one length path in negative ties, so closeness is difficult to compute for node centrality [31].

All the three centralities are three different views on structural properties of graph inferred by Freeman [41].

2.3.4. Eigenvector centrality

This centrality measure was proposed by Bonacich [60]. According to him, centrality or popularity score of a node is proportional to the sum of centralities of nodes to which it is connected. It corresponds to the eigenvector of adjacency matrix W of network:

$$(W - \lambda I)S = 0 \qquad (16)$$

where λ is the eigenvalue and S is the eigenvector. For finding appropriate scores, λ should be the largest eigenvalue of the matrix. This measure was defined for only symmetric matrices and for relations such as friendship that is naturally symmetric. Matrix W should contain value between 0 and 1. If two persons are friends then $W_{ij} = 1$, else 0. For exploring different measures of centrality three approaches were given by Bonacich which explained significance of using eigenvector as centrality measure [60,61].

i. Factor analysis approach:
It suggests that status scores of each actor in network can be calculated by column vector S called the first principal component factor of W, which is also the eigenvector corresponding to the largest eigenvalue of matrix W. The largest eigenvalue is standardized because it represents the length of eigenvector. By factor analysis, clique structure in network can be identified.

Clique is the group in the network in which every member is related directly or indirectly to other members of same group and the relations are not extended outside the group. The factor corresponding to each clique has zero or positive values for their elements. Zero value is for non-member of clique and positive value for member of clique. The individual having high value in factor is the one who is most popular in their cliques. The amount of variance among different factors for each clique depends upon the magnitude of eigenvalue corresponding to eigenvector of popularity scores i.e. larger the eigenvalue, better the eigenvector for analyzing relationships in the clique structure and vice-versa.

ii. Convergence of infinite sequence of status scores:

The simple measure for calculating centrality or popularity, as in case of degree based measure, is the number of friends that a person has. It is called first order measure. This process can be extended by multiplying each person's choice to select others as friends with number of friends others have and is termed as second order measure. It can be improved further by multiplying each person's choice with second order measure. It is named as third order measure and so on. Bonacich [60] expressed this sequence as

$$S_1 = WS_0 \text{ as first order measure where } S_0$$
$$\text{is column vector of ones,} \qquad (17)$$

$$S_2 = WS_1 = W^2 S_0 \text{ as second order measure,} \qquad (18)$$

$$S_3 = WS_2 = W^2 S_1 = W^3 S_0 \text{ as third order,} \qquad (19)$$

and therefore, $S_m = W^m S_0$ as m^{th}order. $\qquad (20)$

This sequence tends to infinity. So to converge it, small modification was made i.e. each measure was divided by the λ_1, largest eigenvalue of matrix W:

$$S_1 = (W/\lambda_1)S_0 \qquad (21)$$

Hence, $S_m = (W^m/\lambda_1^m)S_0$ $\qquad (22)$

He proved that S_m converges to eigenvector of λ_1. This result is same as the output of factor analysis approach. But S_m is converged to give popularity scores of only the largest clique in the network. The smaller cliques and isolates, who are non-members of clique, are scored zero.

iii. Solution of linear equations

In order to find popularity of each person in the network, Bonacich weighted other person's contribution to it by their popularity scores. He generated system of linear equations for unknown scores as [60]

$$S_i = W_{i1}S_1 + W_{i2}S_2 + W_{i3}S_3 + \ldots.. \qquad (23)$$

This was generalized as $S = WS$ in matrix form which gave $(W - I)S = 0$. For finding solution of linear equation, determinant of $|W - I| = 0$. Bonacich made slight modification that did not affect the model but provided a solution to equation as

$$\lambda S_i = W_{i1}S_1 + W_{i2}S_2 + W_{i3}S_3 + \ldots.. \qquad (24)$$

In matrix form, $\lambda S = WS$. It can be written as $(W - \lambda I) S = 0$. This is similar to eigenvector (finding) problem where λ is largest eigenvalue of W and S is eigenvector.

All the three approaches of identifying different aspects of weighted popularity score of actors in network gave the same solution. On this basis, the eigenvector of largest eigenvalue can be considered as a measure of centrality or popularity which is the weighted sum of direct as well as indirect paths of any length to a node in network. This measure is also known as influence measure in which a node influences all of its neighbors and they further influence other nodes. From Table 1, eigenvector allows traffic to move through unrestricted walks as it counts the walks of all lengths originating from a node. It is also appropriate for measuring the long term direct and indirect risk of getting infected in virus network [50] and is used in parallel duplication processes such as attitude influencing [51]. This measure is same as degree based measures but only difference is that degree measures account for walk of one length and in eigenvector walk can be of infinite length. Eigenvector centrality is usually used in networks where degree of nodes varies widely such as low degree node is connected to high degree node. Due to this frequent range

of degrees of nodes, eigenvector can calculate relative score of each node with respect to its neighboring nodes. In networks where degree of nodes does not vary, eigenvector is as good as vector with elements representing each node's degree. Degree based measures are able to work only on simple graph with binary relations between vertices whereas eigenvector centrality measures are used on signed and valued graphs [62]. The variety of complex flows can be analyzed with eigenvector centrality. There are many variants of measures based on eigenvector that accounts for its usage for analyzing different indices. For example, status measure can be calculated in positive and negative networks by forming cliques based on values of eigenvector of adjacency matrix [63] and power can be assessed in negatively connected bargaining networks by using concept of eigenvector [64]. Therefore, this measure can suitably be applied on negative and positive ties network.

3. Influence measures

Besides centrality measures discussed yet, the ties or relations in social network can also be analyzed on the basis of degree of influence, i.e. how much node A gets affected or influenced through relations with node B in network. This can be computed by considering all the connections between them in the network. Unlike the previous measures, that determine the centrality of nodes through geodesics (i.e. shortest paths) only, influence measures consider total number of walks between pair of actors. These types of measures provide relative power or status of a person in a network by accounting all length paths between pair of nodes. But different length paths (such as, 1-length path and 10-length path) cannot be weighted equally for computing influence i.e. a weight depending on length of path should be assigned to each link so that shorter paths get more weights and longer length paths weighted considerably less. On basis of this weight, three different measures were developed: Katz [65], Hubbell measure [66] and Bonacich power measure [64].

3.1. Katz measure

Before 1953, all status measures were considering only number of choices received by an actor as an index of calculating status. However, it provides idea about the high status nodes in network, but correct ordering could not be decided alone on the basis of number of choices or links received by a node. Katz [65] then proposed a new status/influence measure in view to allow not only direct links received by an individual but also popularity or status of individuals sending links to him to be included in his score. Further, the status of each, who has link with these individuals in turn, should also be used for calculating scores in social network. If social network data are represented by a matrix W, also known as choice matrix, then value of each element of matrix W_{ij} represent actor i chooses actor j with condition applicable that W should be positive symmetric matrix [67]. W matrix contains one length path between nodes as its elements whereas power n of W has the number of n-length path chains between pair of actors as elements. This means $W_{ij}^2 = \sum_{m=1}^n W_{im}W_{mj}$ where each component $W_{im}W_{mj} = 1$ only if i chooses m and m chooses j, forming a two length path and W_{ij}^2 element of matrix contain

number of such two length paths [68]. The sum of all the elements of column j in W gives the number of direct choices of j by other individuals. Similarly, column sums of W^2 give two-step choices, W^3 gives three step choices of individual j and so on. Katz defined a measure of status by adding the direct, two-step, three-step etc. choices of individual j by giving lesser weights to increasing order of choices so that probability of longer chains participation be less in computation of status score. Let α be the attenuation factor of W that is the weight assigned to a single link or edge. W contains 1-length path or single links; therefore, its attenuation is α. Similarly, W^2 contains two consecutive links, so its attenuation is $\alpha^*\alpha$, weight of W^3 is α^3 and so on. The end matrix contains sum of all length paths as derived by Katz [65]:

$$T = \alpha W + \alpha^2 W^2 + \alpha^3 W^3 + \ldots + \alpha^k W^k = (I - \alpha W)^{-1} \qquad (25)$$

T is the resultant matrix, attenuation factor 'α' taken in this case, should be such that $\lambda_1 < 1/\alpha < 2\lambda_1$ where λ_1 is absolute magnitude of largest eigenvalue of matrix W and its value depends upon the group of individuals to which information is transmitted. Through Katz measure, most influential node or individual in positive tie network can be found who has connections with most of the other individuals and can influence or affect other individuals with his decisions or activities. The major conclusion drawn by Katz was that the status of individual in network depends not only on number of direct links but also on the status of next to direct links. For example, it can be possible that nodes B and D having 5 and 4 direct links are of low status as compared to node A which has only two immediate links with B and D. This happens because A is connected to two nodes that are high status nodes in network and node A influences all other nodes in network through these two nodes and thus obtains highest status in the network, whereas B and D get comparatively less score. As shown in Table 2, Katz measure can be applied in parallel duplication flow processes and generally follows walks of infinity length [51]. This index was not interpreted to be applied on network that has negative relations as it depends upon flow in the network and its value depends upon how many well-connected nodes are immediately connected to a node. It is concluded by Everett and Borgatti [31] that "In negative ties the value of measure that flow through networks is analyzed rather than a flow of resources as opposed to positive ties network in which value of something flowing through network is analyzed through centrality measures". But Katz only tried to capture the value of something flowing through network of symmetric relations such as friendships.

3.2. Hubbell measure

Hubbell [66] interpreted the flow of influence through interpersonal links in social networks as input and output channels. He analyzed the association of dyads or pair of nodes in network for identification of cliques as well as measuring the contribution to the status of individual by other persons present in the network that is through computing dyadic influence. The index proposed by Hubbell has structural as well as functional significance. The structural significance of index is in identifying cliques and functional significance is in computation of status scores. He considered the structure matrix W (same as choice matrix W in Katz measure) in which $W_{ij} = 1$ if j chooses i

Table 2 Characteristics of three influence measures.

Type of influence measure	Katz influence	Hubbell influence	Bonacich power
Proposed by	L. Katz	C. H. Hubbell	P. Bonacich
Year	1953	1965	1987
Expression of status scores	$t = ((I - aC) - I)^{-1}$	$h = (I - \beta W)^{-1}E$	$b = (I - \beta R)^{-1}R1$
Attenuation factor	$1/a < \lambda_1$ where λ_1 is the characteristic root of C	β is the fractional weight assigned to entries of matrix	$\|\beta\| < 1/\lambda_1$ λ_1 is the largest eigenvalue of matrix R
Type of graph (symmetric/ asymmetric)	Only symmetric graphs	Only symmetric graphs	Both for symmetric and asymmetric graphs
Type of network ties supported	Interpreted only for positive ties network	Interpreted for both positive and negative ties (by varying signs of value of β)	Both for positive and negative ties networks
Relation between three measures	$t = h - 1$ $t = \beta b$ where t is Katz measure, h is Hubbell measure and b is Bonacich measure		
Matrix used in measure	Choice matrix C, $C_{ij} = 1$ depicting actor i choosing actor j	Structure matrix W, W_{ij} depicting weights of links between nodes that can be positive, negative, zero or fractional	Relationship matrix R, R_{ij} depicting values between 0 and 1 inclusive

and W_{ij} is the weight of link joining i with j. The value of weight can be positive, negative or zero. All the powers of structure matrix are used to consider the entire one to n-length chains between pair of nodes. Hubbell [66] performed summation of all powers of matrix and called it Y:

$$Y = I + W + W^2 + W^3 + \ldots \qquad (26)$$

This infinite series can be converged by having fractional weights in W structure matrix. As there exists two values for each dyad y_{ij} and y_{ji}, index of association of dyads is defined as $m_{ij} = m_{ji} = \min(y_{ij}, y_{ji})$ and used for identifying cliques by discriminating inter-clique and intra-clique bounds. m_{ij} and m_{ji} are the structural significance of index. Functional significance was described by status score vector s. Let s_i be the status score of ith member of group structure and is given as [66]

$$s_i = e_i + w_{i1}s_1 + w_{i2}s_2 + w_{i3}s_3 + \ldots \qquad (27)$$

where e_i is the exogenous factor and w_{ij} is weight of link i to j. In matrix form

$$S = E + WS = (I - W)^{-1}E \qquad (28)$$

W is the structure matrix. Here, values of column vector E are also known as boundary conditions for a system. It became necessary to include E when open systems such as communication networks are considered. Status score column vector S is defined by Hubbell as [66]

$$S = (I + W + W^2 + W^3 \ldots)E \qquad (29)$$

The interpretation of Hubbell measure is slightly different from Katz in which w_{ij} represent weight of influence of j on i whereas in Katz w_{ij} represent the weight by which i chooses j. Katz also ignored the boundary conditions in open system as presence of E in Hubbell's measure. The attenuation factor in Hubbell measure is the fractional weights of structure matrix W. These weights converge infinite series of powers of W. Katz did not include identity matrix in series for calculation of total number of walks between pair of nodes. Both measures computed the flow of influence in network. In Hubbell's structure matrix, negative values of weight of links are allowed; therefore, it can be used to analyze both the positive and negative relations in sociometric data. Hubbell measure can be used in parallel duplication flow processes such as attitudinal influencing [51].

3.3. Bonacich power measure

The centrality of node does not always correspond to power of node in social networks [69]. Specifically in negatively connected bargaining networks, the most central person is not always the most powerful one. This view contradicts the extensive social literature [70–72]. In order to distinguish between power and centrality, Bonacich [64] proposed a set of measures given by $c(\alpha, \beta)$. The parameter β is used to reflect the degree and direction (positive or negative) in which individual status depends upon status of other people in network. In positive communication network of information exchange, power of a person depends positively on high status persons he/she is connected to. The parameter β attains positive value in these networks. But in case of negatively connected networks of commodity exchange, power of a person decreases by having connection to high status people who have many potential alternatives. In such situation β has negative value. The magnitude of β reflects the degree to which status score $c(\alpha, \beta)$ of a person is a local or global measure i.e. to which extent it is a function of status score of immediate other persons in network. If $\beta = 0$, then $c(\alpha, \beta)$ is similar to number of direct communication links to all other persons in network. As β increases, centralities of distant persons or indirect communication links are considered in $c(\alpha, \beta)$ score. This measure can be applied to symmetric as well as asymmetric networks [73].

This measure is based on eigenvector centrality 'e' proposed by Bonacich, given as [60]

$$\lambda e_i = \sum_{j=1}^{n} A_{ij} e_j \qquad (30)$$

where A is adjacency matrix and λ is the largest eigenvector of matrix A. In the same spirit, beta centrality is defined as [64]

$$c_i(\alpha, \beta) = \sum_{j=1}^{n} (\alpha + \beta c_j) R_{ij} \qquad (31)$$

Now, centrality of neighbor nodes is weighted by parameter β whose magnitude and sign define the contribution of other actors of network to power of actor 'i'.

In matrix form, the measure is given as [64]

$$c(\alpha, \beta) = \alpha(I - \beta R)^{-1} R1 \qquad (32)$$

$$c(\alpha, \beta) = \alpha(R1 + \beta R^2 1 + \beta^2 R^3 1 + \ldots) \qquad (33)$$

This is the series of path-length chains. Here, absolute value of β is less than reciprocal of largest eigenvalue of R. This expression indicates the contribution of direct links and all length paths (indirect links) of each individual in the beta centrality measure. Katz centrality closely resembles the beta centrality in estimation of status of an individual in network. Let t be the Katz centrality, given as

$$t = \beta c(\alpha, \beta) \qquad (34)$$

When β is positive both measures are perfectly correlated but when β is negative Katz and Bonacich measures become negatively correlated [74]. β in the expression can attain two values either positive, zero or negative. The measurement of power in both cases is as follows:

i. When β is positive or zero:

If $\beta = 0$, it means $c(\alpha, \beta)$ is simply a degree measure and status of person is defined by number of people directly connected to it. But when β is non-zero and positive, magnitude of β defines the degree to which score of actor is a function of scores of other actors in network. The positive value of beta is achieved in positive network, such as communication network, where information is exchanged between pair of actors. β is the probability that information is received by actor and all of its contacts. According to Eq. (33), $R1$ is the direct information exchange to connected ones, $\beta R^2 1$ is the transmission of information to actors over 2-length paths and so on. The value of β decides up to which level of path length chains the information should be delivered by an actor to surrounding actors. It can be assumed that a circle is defined by radius β and value of β decides to exchange information either locally or to the structure as whole. In communication networks, the higher value of β would be appropriate to communicate over distant nodes. The interpretation of β in positive network is similar to Katz measure i.e. power of person in social network depends upon status of people to whom he/she is connected. Higher the status of surrounding people, more powerful the person in network.

ii. When β is negative:

β usually attains negative value in negative networks, such as bargaining and exchange networks. Here the interpretation of power completely changes. Suppose in exchange network, commodity exchanged with one person cannot be exchanged with another and bargaining power of person decreases if he/she is connected to high status people who have many other potential alternatives.

Reconsidering Eq. (33), the function of β can be analyzed as follows:

If β is negative in above equation, then terms of even power of R decrease the score of person and terms of odd powers of R increase value of $c(\alpha, \beta)$. This can be interpreted as follows:

a. $\alpha R1$ indicates that having many connections increases power and centrality.
b. $\alpha\beta R^2 1$ implies that if the connections of person have many other alternatives in exchange networks then it decreases power of person but increases his/her centrality.
c. $\alpha\beta^2 R^3 1$ means if there are many paths of 3-length or many connections of nodes at 2-length path then it decreases power of immediate connections of person and consequently increases his own power [64].

This can be explained by the following example:

In Fig. 8 of tree graph depicting exchange network, the power of A increases due to connections to B and C at length one. But, as there are many connections to B and C at 2-length which make them central, it results in decrease of power of A. Consequently, D, E, G, and H are also having many connections, which decreases power of B and C but increases the power of A. This concludes that power of A is increased due to more number of nodes at 1-length and 3-length paths and decreased due to the presence of many nodes at 2-length paths.

Thus, beta centrality differs from Katz measure of calculation of status. $c(\alpha, \beta)$ measure is useful in valued and signed graphs, negative ties and positive ties networks. It is especially sensitive in situations where many low degree nodes are connected to high degree nodes i.e. where difference of degree is present to drive centrality. In many situations, beta centrality is similar to degree measure, such as in regular graphs where degree of nodes does not vary when elements of eigenvector corresponding to largest eigenvalue sum up to zero and when there are multiple eigenvalues of same magnitude existing for symmetric matrix [62].
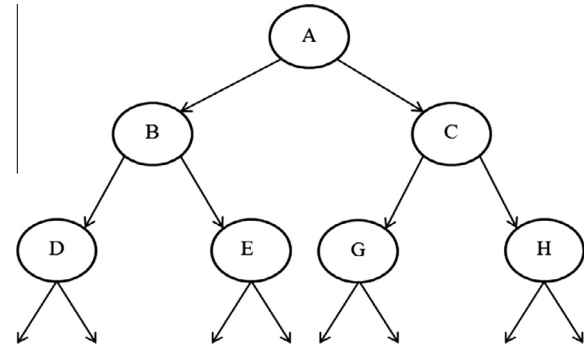
Figure 8 Exchange Network, where A is focal actor having connections with B and C with whom he exchanges commodity.

4. Projected approaches for analyzing ties

### 4.1. Graph complement approach

As all standard approaches and centrality measures are not suitable for negative ties but applicable to positive relations in network, in order to implement these techniques on negative networks, graph complement approach can be used [31]. In this approach, graph containing negative ties is complemented to form positive tie network. In complement graph, all edges that were present in negative tie graph are absent and vice-versa. Now this graph represents positive tie network and all network analysis techniques can be applied to this graph. This is very intuitive approach and can make its way to apply all techniques in negative tie data. For example, consider the Fig. 9 depicting the negative ties graph (G) and its complement (G') containing edges that are absent in G.

There can arise many problems with this approach. First, it is not necessary that absence of negative ties between nodes means positive relations in complement graph. If the network states that a node exists in one of the two states, such that if a relationship is not in state 1 then it should be in state 2 only, then the graph complement approach can work out. Second, the property of graph states that both G and G' cannot be disconnected simultaneously i.e. if graph of negative ties G is connected then it is not necessary that G' is also connected. In some situations, G' becomes disconnected, then centrality measures such as closeness and betweenness cannot be used for analyzing ties. Third, besides structural differences of properties of graph and its complement, only degree centrality can be analyzed in G'. Suppose, the degree of node 'a' in G is x, then degree of same node in G' becomes $n - 1 - x$, where n is total number of nodes in G. Degree centrality in G' is derived from degree centrality in G. But, it is not true for betweenness and eigenvector centralities. The methods for which results in complement graph can be deduced from original graph are called as complement consistent methods [31]. To find equivalent actors, the approaches such as structural equivalence and automorphic equivalence are complement consistent but regular equivalent nodes are not present in complement graph. In fact, all centrality measures are not complement consistent except degree centrality. The density of negative tie network is usually low; thus, complement graph G' has high density and centrality values show little variance. All the methods discussed in previous sections that are interpretable for both

positive and negative networks yield results both in G and G', but measures that are not applicable in negative networks due to their structural features can be analyzed through complement graph [31]. As there are many problems with complement approach also, there is a need of new set of measures that can be applied specifically on negative tie data and will be discussed in next section.

4.2. Negative cohesive subgroups

There is a vast literature on concept of cohesive subgroups in networks that are formed as a result of relationships between the nodes. This subgroup is popularly known as 'clique'. Luce and Perry [75] defined clique as "the subset of group consisting three or more members, each having symmetric relationship with each other and no member outside the subset exist that is in symmetric relation to every member of subset". Many researchers have defined clique and also developed many measures for identifying them in networks. For example, Hubbell [66] in 1965 developed an input–output approach for clique identification, Bonacich [60] in 1972 developed eigenvector centrality measure for finding cliques in symmetric graphs of social network and Everett and Borgatti [76] in 1998 analyzed the overlapping cliques in network. The structure of clique can be recognized from various powers of symmetric matrix, say A (derived from adjacency matrix). The main diagonal entries of matrix A^2 ($a_{ii}^{(2)}$) denote the number of chains of length two and also the number of elements in graph with which element 'i' is in symmetric relation. This set of elements along with i forms a clique. In order to identify whether a given element belongs to clique, its main diagonal entry in A^3 is checked for positive value. If this entry is positive then element belongs to clique and if it is zero then element is not a member of any clique [75]. The complete structure of clique was not defined by them. Everett and Borgatti [76] tried to determine the overlapping among cliques in network by analyzing their structures. They made use of UCINET tool of social network analysis [77]. The standard method of clustering was used by the tool to identify groups in network but failed to identify structure and overlapping of cliques. Therefore, they defined clique graphs and clique overlap graphs. Clique graph or intersection graph has clique as vertices and edge between them represent that cliques have some actors in common. Clique overlap graphs are iterative clique graphs that represent the amount of overlapping between cliques in each iteration [76]. For

Graph 'G' of negative ties   Complement Graph G' of positive ties
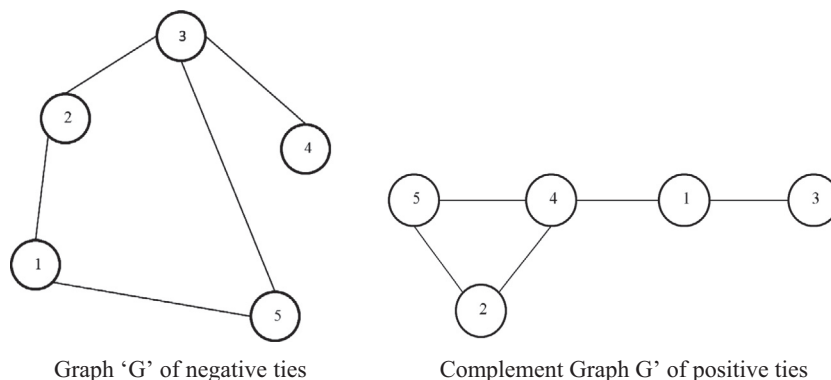
Figure 9   Graph of negative ties network and its complement depicting positive ties network.

example, there are six cliques namely 1, 2, 3, 4, 5 and 6 as shown in Fig. 10, each having actors common among them as depicted by numbers on edges. This graph G is named as clique graph, $K(G)$. The iterative overlap graphs are $K^2(G)$, $K^3(G)$ and so on. $K^2(G)$ contain edges between cliques which have two or more than two actors in common and similarly, $K^3(G)$ contain edges only between cliques having three or more than three actors in common. The clique overlap graphs describe amount of overlap among cliques and also help in analyzing structures of cliques. By using these graphs, Everett and Borgatti [76] defined co-clique matrix C, with entry c_{ij} as the number of actors that clique i has in common with clique j. The results of implementation showed that structures and overlapping of cliques were clearly identified. The entries of actor by clique matrix contain the number of cliques in which each actor is present. This is known as clique overlap centrality. When this measure was compared with other centrality measures, results showed that it correlates with others in core or periphery regions of network. When multiple such regions are present then this measure gives different values as output [76]. The above discussion proves that clique is very important concept in analyzing positive ties in network. But in case of negative ties, clique represents a group in which each member is in negative relation with each other and group formed is also not cohesive. Everett and Borgatti [31] defined negative clique in which every member outside the group has negative relations with members of the group. This group is the smallest group of individuals in network such that everyone outside the group dislikes at least one individual from inside the group. The negative cliques also overlap with each other and this overlapping can be analyzed with same methods that were used for ordinary cliques. The negative clique can be generalized by specifying its maximum size (just as minimum size in general clique) and defining negative k-plex clique. The negative clique overlap centrality can be computed for each actor by counting number of negative cliques in which it is present. Negative clique usually represents the highly undesired, disruptive group of individuals and thus disliked by everyone outside the group. Negative cliques can be found by key player program [78] in which those key players form a group, who can reach everyone in network. This negative clique concept is future area of research for analyzing negative ties. It can

be used in identifying the group of pupils in classroom who are responsible for bullying other students. Such a group is disliked by everyone in the class and forms negative cliques.

5. New measures of negative ties

Earlier approach of graph complement requires some new measures that specifically deal with negative relations and helps in identifying such ties between pair of actors in the network. Therefore, existing measures of positive relations such as degree centrality and eigenvector centrality are tailored in such a way so that they satisfy conditions of negative ties network. These new measures are given by Everett and Borgatti [31] and are analyzed by applying on directed networks.

5.1. Negative degree centrality

Degree centrality of a node in negative networks is defined as degree of a node in complement graph, $n - 1 - x$, where x is degree of node in graph G [31]. In order to normalize, divide this degree by $n - 1$, i.e.

$$d^*(x) = 1 - x/(n-1) \tag{35}$$

This measure gives low score to actors who are disliked by most of the others and high score to actors who are disliked by less people. In directed graphs, d^*_{in} and d^*_{out} are the in and out degrees of nodes. High d^*_{in} score means that an actor receives less number of negative ties and high d^*_{out} score means that an actor sends lesser number of negative ties to other actors in network. A person with low d^*_{in} and d^*_{out} score is the most disliked person and similarly an actor is most popular in network if he/she has high d^*_{in} and d^*_{out} score [31]. Negative degree weights every actor equally and does not consider the position of actor with respect to other popular people in network i.e. an actor connected to highly central actor (most popular actor) gets relatively high score as compared to an actor connected to disliked person in network. This measure is related to structural property of node which signifies the number of nodes it is connected to but not to which node it is connected. It fails to identify the centralities of neighbor nodes and their effects on centrality of given node [31].
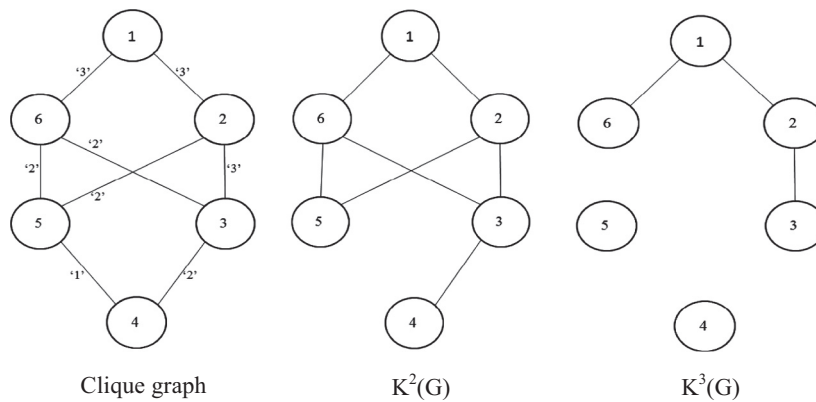
Figure 10 Clique graph, first iterative overlap graph $K^2(G)$ and second iterative $K^3(G)$ graph. Nodes of graph depicting cliques and number of overlapping actors are written as labels of edges.

5.2. h^* centrality

As negative degree measure has its limitation in evaluating centrality of node, there is a need of measure that considers negative connections to less central (most disliked) person is better rather than having negative ties with most central (popular) actor in network. This aspect of calculating centrality is similar to eigenvector centrality as we discussed in Section 2.3.4. Everett and Borgatti [31] have recently proposed a new measure known as h^*. The centrality of node x is denoted by $h^*(x)$ [31]:

$$h^*(x) = 1 - \sum_{y \in N(x)} \frac{h^*(y)}{n-1} \qquad (36)$$

where $N(x)$ represents neighborhood of node x. In matrix form, equation is given as

$$h^* = \left(I + \frac{1}{n-1}A\right)^{-1} 1 \qquad (37)$$

where I is identity matrix, 1 is a column vector of 1's and A is the adjacency matrix. This measure is similar to Katz [65], Hubbell [66] and Bonacich [64] measures. $h^*(y)$ is similar to eigenvector centrality of node y. Only difference is that Eq. (36) is normalized by dividing it by $n-1$ and to bring it in 0 to 1 range it is subtracted from 1. The h of Eq. (37) is similar to Hubbell measure, $h = (I - \beta A)^{-1} 1$, where $\beta = -1/(n-1)$. But Katz and Hubbell did not consider negative beta. The concept of negative beta was proposed by Bonacich [64] in negatively connected bargaining networks where it is more beneficial to connect with ones that have less options or alternatives. This increases centrality of node. The same implication is used in h^* measure for negative tie networks and used negative beta. This measure has time complexity of $O(n^3)$. In case of directed networks, h^*_{in} and h^*_{out} are also calculated for each node. An actor with low h^*_{in} score receives many negative ties from other actors, especially those who send very less negative ties to other actors and have high h^*_{out} score. This means receiving negative ties from popular actors of network greatly affects the centrality of actor. In similar fashion, an actor with low h^*_{out} score is the one who sends negative ties to most of the actors, especially those who have very less incoming negative ties and have high h^*_{in} score [31]. It means this person dislikes the one, whom everyone likes. The directed formulae of h^* measures are given as follows [31]:

$$h^*_{out}(x) = 1 - \sum_{y \in N_o(x)} \frac{h^*_{in}(y)}{n-1} \qquad (38)$$

$$h^*_{in}(x) = 1 - \sum_{y \in N_i(x)} \frac{h^*_{out}(y)}{n-1} \qquad (39)$$

N_o and N_i are out and in neighborhood of node x. Equations in matrix form are as follows:

$$h^*_{out} = \left(I - \frac{1}{(n-1)^2} A A^T\right)^{-1} \left(I - \frac{1}{(n-1)^2} A^T\right) 1 \qquad (40)$$

$$h^*_{in} = \left(I - \frac{1}{(n-1)^2} A^T A\right)^{-1} \left(I - \frac{1}{(n-1)^2} A\right) 1 \qquad (41)$$

where A^T is the transpose of matrix A. The h^* measure presents more accurate and precise results as compared to negative degree centrality d^*. For each actor in network h^* measure gives unique value of centrality as compared to d^* which gives multiple actors the same score. Therefore, h^* is refinement over d^* [31]. Both h^* and d^* measures can be extended to accommodate valued data only when largest eigenvalue of adjacency matrix is less than $n-1$.

6. Mixed data measures

Today's online social networks such as Facebook, Twitter, Orkut, and LinkedIn include all types of relations of friendship, support, liking, disliking, avoidance, distrust or no relation at all between various nodes. The measures discussed till now are applicable either on positive relations or on negative relations but in order to cater all aspects of relations possible in social networks, there is a need of measures that can analyze both of these relations jointly. Many researchers have proposed measures that consider data of both types.

6.1. Degree centrality measure

This measure is similar to negative and positive degree centrality discussed earlier. Pair of actors connected to each other through positive relations has positive degree and if they are connected negatively then they have negative value of degree with respect to each other. Thus, overall degree of an actor in network is the summation of negative and positive values of degree as shown in Table 3. The person with overall high negative value of degree is the most popular one in network. This measure gives the idea of status of person in network but again this does not generate precise value if two actors are having nearly similar connections with other actors in network. This means it can assign same value of degree to two or more persons in network [31]. Also the effect of popularity of person on status of neighbor nodes is not included in this measure. Therefore, a more general measure is required.

6.2. Status measure

Bonacich and Lloyd [63] proposed status measure using eigenvectors. In a network of both positive and negative relations, a status of individual rises if he/she is positively connected to high status individual and status decreases by having negative relations to the same. If individual is having positive relations with negative status persons then also his/her status reduces. This measure is defined by using eigenvector and diagonal matrices. Adjacency matrix A, has a_{ij} equal to 1 if i and j are positively connected, -1 if they are negatively connected and 0 if there is no relationship. Therefore, it forms a balanced graph. The eigenvector of A corresponding to largest eigenvalue with positive and negative elements reveals the clique structure. All the actors with positive value of status belong to one clique and negative value actors belong to another clique. Positive relations exist between members of same clique and negative relations between members of different cliques. The status of individual actor increases with increase in number of positive connections to members of same clique and number of negative connections to members of different cli-

Table 3 Comparison of Mixed data measures.

Type of mixed data measure	Degree centrality	Status measure	Political Independence Index (PII)	PN centrality
Proposed by	Positive degree by L. C. Freeman Negative degree by Everett and Borgatti	P. Bonacich and P. Lloyd	Smith et al.	M.G. Everett and S.P. Borgatti
Year	1979 2014	2004	2014	2014
Expression	$P(x) + N(x)$, $P(x)$ is number of direct positive ties and $N(x)$ is no of direct positive ties	$Ax = \lambda Dx$ A is adjacency matrix; Dx is eigenvector depicting status scores of each node	$\sum_{i=0}^{k} \beta^i [P(i)^x - N(i)^x]$ $P(i)$ and $N(i)$ are positive and negative edges incident on node	$\text{PN} = \left[I - \frac{1}{2n-2}A\right]^{-1}$ $A = P - 2N$, P is positive tie matrix and N is negative tie matrix
Weightage of direct and indirect links	Direct links = 1, Indirect links = 0	Direct links = 1, Indirect links = β^k, where $\beta = 1/\lambda_l$, λ_l is the largest eigenvalue, k is the number of edges between nodes	Direct links = $P(0)^x - N(0)^x$, Indirect links = $(\beta)^i$, i is the distance of edge from focal actor and x an exponent which ensure that more weight is assigned to direct links as compared to indirect links	Direct links = β, Indirect links = β^k, k is the path-length from given actor, $\beta = 1/(2n-1)$, for large network size $\beta = 1/(2\delta)$
Applicable on types of graphs	Both directed and undirected graphs	Only on undirected graphs because on directed graph eigenvalue may not be real	Only on undirected networks	Both on directed and undirected graphs
Concept used for analyzing ties	Positive and negative relations with nodes connected directly	Based on balance theory forming two cliques where positive relations exist within cliques and negative relations between cliques	Based on concept of positive and negative edges at a distance of 1 from a node in political networks	Based on positive P and negative tie N matrices. Matrix A is formed by: $A = P - 2N$
Complexity	$O(n)$	$O(n^3)$	$O(n^3)$	$O(n^3)$

ques [63]. This measure solves the problem of analyzing positive and negative ties jointly up to some extent. But in some cases, when balanced symmetric adjacency matrix A has multiple large eigenvalues having same magnitude, then there exists multiple linearly independent eigenvectors [31]. Out of these, it becomes difficult to choose which eigenvector should be used to assign status score. Secondly, if network has multiple components or cliques similar in structure, that is connections between members of same clique and members of other clique are similar for multiple cliques, then the eigenvector of matrix A is not able to describe score of all the actors of component correctly [62]. Thus, eigenvalue assigns anomalous scores to actors of the cliques. Also, if graph is directed then eigenvalues may not be real.

6.3. Political Independence Index (PII)

Measures discussed till now, focused on actor's characteristics and attributes in positive networks. All of them counter the direct and indirect access of flows of information between pair of nodes for analyzing positive and negative ties as well as power or status acquired by each actor in network. Smith et al. [79] proposed a power measure, known as Political Independence Index (PII), according to which any actor (also called as focal actor) in the network is dependent on other actors in respect of control of resources, support and

information. Those other actors become more powerful than focal actor in politically charged networks of both positive and negative ties. According to PII measure, an actor can reduce his/her dependence by developing alternatives that are weaker and itself in direct threat. They classify power measures as power-as-access that include degree, closeness, eigenvector, Bonacich-power-measure with positive beta and power-as-control that include betweenness, Emerson's social exchange theory [80,81], Graph theoretic power index (GPI) [82,83], and Bonacich-power-measure with negative beta. All these power-as-control measures compute power of an actor as an extent to which actor controls flow of resources and exchange of information between pair of other actors through direct or indirect links. PII measure is also based upon power-as-control approach and is highly related to Bonacich power measure with negative beta. This index is defined for politically charged networks consisting of alliances (positive ties) and adversaries (negative ties). Smith et al. [79] proposed that when adversaries of an actor grow then actor forms alliances with other friendly actors to safeguard itself in exchange of some resources. The extent to which resources are sacrificed depends upon urgency in making alliances with alternatives. When number of alternatives (allies) is limited for making alliances then these potential alternatives can extract maximum resources from focal actor which decreases its power. These allies must not have many other alternatives with whom they can make alliances as it

directly decreases power of focal actor. The measure uses concept of distance of an edge (m, n) from focal node u as MIN $(d(u, m), d(u, n))$ where d is geodesic distance between nodes. For direct ties, $d(u, u)$ is 0 and $d(u, v)$ is 1. Therefore, distance of edge (u, v) from u is 0 and progressively it increases by 1 at each length path.

PII measure can be stated as [79]

$$\sum_{i=0}^{k} \beta^i [P(i)^x - N(i)^x] \tag{42}$$

where $P(i)$ and $N(i)$ are the number of positive and negative edges at distance i from a node, β is the attenuation factor and x is an exponent defined as

$$x \leq \frac{\ln(2) - \ln(|\beta|)}{\ln(M)} \text{ i.e. } \beta M^x \leq 2 \tag{43}$$

Here, M is the maximum number of ties from any node in network. The attenuation factor provides flexibility to measure by behaving both as power-as-access measure when β is positive and power-as-control when β is negative. As β has power of i, which is the distance of edge from a node, it takes into account the fact that in the presence of more number of alternatives or allies to direct link node at distance one decreases the power of focal node. The exponent factor x is used to take into account the contribution of two direct links to power of focal actor, more than one direct link and one indirect link. This implies that nearby actors in network greatly affect power of focal actor as compared to distant actors [79]. The interpretation of this measure is slightly difficult in mixed network of positive and negative ties. For example, consider network of mixed data in Fig. 11.

According to status measure, actor B is given high score as it is connected to other well-connected actors such as C, D and A. But PII measure assigns very low score to actor B as it is least powerful because it has direct links with C, D, A who are having allies or alternatives E, F, G, H that are totally dependent on them with no positive connections to any other nodes except C, D, A. This example shows that PII does not capture the idea of centrality like other normal centrality measures do. It actually deals with power of an actor which
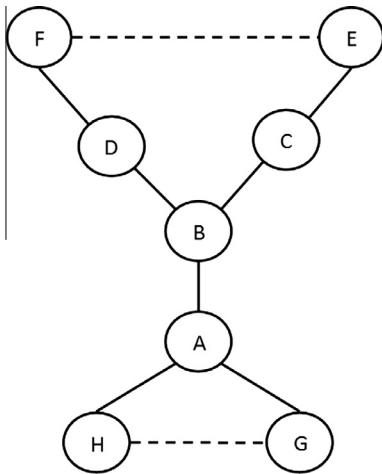
Figure 11 Network of nodes having positive relations depicted by solid lines and negative relations depicted by dashed lines.

gets differentiated from centrality when negative ties are considered.

6.4. PN centrality measure

Fourth measure is an extension of h^* measure in negative data networks. Everett and Borgatti [31] proposed PN centrality measure which uses h^* measure for negative ties and Hubbell measure for positive ties. The value of β is chosen as $1/(2n - 2)$ to normalize Hubbell centrality scores in complete network and matrix A is calculated by $P - 2N$, where P is positive tie matrix and N is negative tie matrix [31]:

$$PN = \left[I - \frac{1}{2n-2}A\right]^{-1} \tag{44}$$

Directed version of PN measure is

$$PN_{in} = \left(I - \frac{1}{4(n-1)^2}A^T A\right)^{-1} \left(I + \frac{1}{2(n-1)}A^T\right) \tag{45}$$

$$PN_{out} = \left(I - \frac{1}{4(n-1)^2}A A^T\right)^{-1} \left(I + \frac{1}{2(n-1)}A\right)1 \tag{46}$$

The values of PN_{in} and PN_{out} are analyzed in the same way as h_{in}^* and h_{out}^*. Lower the value of PN_{in} means that person is having more number of incoming negative ties from other people, particularly from those who are popular in network i.e. having high PN_{out} score people. Similarly, high PN_{in} value means receiving lesser number of negative ties and more number of positive ties. Various other permutations are also possible. Range for PN scores for only positive tie network is 1–2, for only negative tie network it is 0–1 and for both positive and negative ties network it is −1 to 2. Complexity of PN is $O(n^3)$. Other values of β can be used such as $1/(2\delta)$, where δ is maximum positive or negative degree of any node in network. But this value of β should always be smaller than $1/\lambda_l$, where λ_l is the largest eigenvalue so that centrality scores of actors should remain proportional to scores of neighboring actors [31].

Negative tie measures discussed yet, are based upon the scores of neighboring actors to which a given node is connected. In such measures, only the scores of actors propagate through network as it is very difficult to interpret flow of resources through it. As opposed to positive ties, where centrality measures try to capture flow of information, resources, etc. through network due to high level of transitivity, the value of measure does not need to propagate through it. The measures that are used for analyzing negative ties are all based on this idea of capturing flow of scores assigned by indices.

7. Strengths and weaknesses of measures

The strengths and weaknesses of all the measures discussed in this paper are summarized in Table 4.

As it is clear from the table, only mixed data measures are efficient in analyzing both the positive and negative ties simultaneously in social network, whereas, centrality and influence measures are able to identify only one type of tie at a time (i.e. either positive or negative). Everett and Borgatti [31] analyzed the accuracy of mixed data measures in identifying outsiders of social networks by applying them on small datasets

Table 4 Strength and weaknesses of all measures.

Measure types	Measure name	Strength	Weaknesses
Centrality measures	Degree centrality	• It is independent of flows in network • It is used to calculate the direct influence • It can be used for analyzing both the negative and positive ties	• It cannot calculate the centrality possessed by a node due to the presence of nodes, located at more than one length path from a given node
	Betweenness centrality	• It calculates centrality of node based upon its property to control stream of information passing through network • It can be used for identifying positive ties	• It is dependent on flows in network • It cannot analyze negative ties network where flows are absent
	Closeness centrality	• The centrality score given by closeness computes node's efficiency in delivering message to other nodes of network in shortest possible time • It can identify all the connections that a node has in positive networks	• This measure is totally dependent on flows in network • In case of disconnected graphs of negative networks, closeness becomes zero for unreachable nodes
	Eigenvector centrality	• It calculates centrality possessed by a node based on centrality of neighbor nodes located at one or more than one length path • It also includes the impact of centrality of neighboring nodes in score of each node • It can be applied on both positive and negative networks	• In case of very large networks, eigenvalues of network may not be real or there may exist multiple eigenvectors
Influence measures	Katz measure	• It was observed first by Katz that the status score of node depends not only upon direct links but also on status of next to immediate links • It works perfectly well with positive ties network	• It is not suitable for analyzing negative ties in a network as it depends upon flow of resources
	Hubbell measure	• It allows negative weights of links, therefore, includes negative relations, such as enemies, in computation	• It is not effective in calculating centrality of relations in negative networks
	Bonacich power	• It computes centrality of nodes based on all walk lengths and can be used in both positive and negative networks	• Though it is able to analyze ties in both networks by using positive and negative values of β but cannot analyze both the ties simultaneously in mixed networks
Mixed data measures	Degree measure	• It can identify both the positive and negative ties simultaneously in mixed networks	• It does not consider the effect of popularity of person on status of neighbor nodes • It fails to distinguish between two actors if they have similar connections
	Status measure	• It can identify both the positive and negative ties in small networks	• In larger networks multiple larger eigenvalues may exist with same magnitude • Also there may not exist any real eigenvalue which becomes difficult to choose single eigenvector that will represent correct status scores
	PII measure	• It correctly analyzes power of node in network of allies and adversaries	• It cannot identify the most positive and most negative nodes in correct order in mixed data networks
	PN centrality	• It can identify the popular and unpopular nodes of network correctly with unique score assigned to every node	• The results of this measure are still unexplored in large online datasets

such as Sampson's monastery [17] and Bank Wiring Room [15]. These datasets were gathered by original researchers physically present at location, inferring the relationship between nodes by observing their behaviors. They pointed out that PN centrality is better than status and PII measures as both of them are not able to identify either central or peripherals (isolates) in the network. When PN is compared to degree, they concluded that PN is more accurate and maps small differences between centralities of nodes to considerably precise values that can be easily compared for analysis purpose. The accuracy of result obtained by Everett and Borgatti [31] was determined merely on the basis of analysis made from application of measures on small offline datasets. The results could have been verified by applying on large datasets of current online social networks with thousands of nodes interacting with each other. For verifying the same, we have applied these measures on sample collected from large online network of Epinions and Slashdot and observed that PN centrality outperforms the other measures. The result obtained indicates that only PN measure identifies most of the outsiders of network efficiently as compared to degree measure which is unable to identify the most popular nodes of network correctly. The status measure becomes inapplicable on large datasets due to imaginary eigenvalues of matrix. Also, PII measure is not able to identify outsiders in large datasets and therefore, performs similar to that observed in small datasets. Further

the work can be carried out to enhance the performance of PN centrality in datasets of online social networks by varying the value of parameter β.

8. Conclusion and future scope

The exploration of negative ties in social networks becomes very necessary in today's scenario where social interactions are carried out online and reality of intended person can be fake with whom one communicates. This may lead to many social crimes such as masquerading, bullying, and group conflicts which need to be dealt with by observing the structure and connections in network. This study of analyzing networks is known as social network analysis. The analysis of negative ties such as distrust and opposition may also help many users of consumer review websites such as Epinions.com and Slashdot.com in deciding which product to purchase by checking their rated reviews. In this paper, a survey of network analysis techniques that can be used to examine such type of negative relations is presented. From discussion, it has been observed that negative ties face difficulties in exploration through standard methods and techniques. The view of analyzing negative ties, in the same way as positive ties, merely by reversing the interpretation is not completely true.

The paper is structured into different sections. Section 2 discusses the application of standard concepts such as equivalence of nodes, statistical models, and centrality methods on negative tie networks. Some concepts are perfectly suitable such as equivalence concept and statistical techniques while many others such as degree centrality require few alterations in interpretation but are applicable in both ties. Other concepts such as betweenness and closeness centralities rely on network flows and thus cannot be applied on negative ties where flow among nodes of a network is minimum. Difference between centrality measures is presumed on the basis of various parameters and type of flows they follow. In third section, examination of influence measures applied to negative ties shows that some measures are easily adaptable to positive and negative ties by adjusting value of attenuation factor, such as Bonacich power and Hubbell measures, while others are inapplicable due to absence of network flows in negative ties.

Due to these issues, some projected approaches, such as Graph complement approach and Negative cliques, that can allow positive tie measures to be applied on negative ties data by reversing the interpretation are discussed in fourth section of paper. But they too pose some restrictions on type of networks. Therefore, some new measures such as negative degree and h^* measure were developed that were designed specifically for analyzing negative ties. These are discussed in Section 5 of the paper. Out of these two, h^* measure proves to be more efficient as it is based upon concept of calculating centrality of given node by considering the centralities of neighbor nodes. In order to analyze both types of ties jointly, mixed data measures proposed by many researchers are described in Section 6 and comparisons are made in terms of their capability in identifying negative ties in network. The strengths and weaknesses of all the discussed measures are presented in seventh section. From this discussion, it is observed that only PN centrality measure is able to work out in practical social networks and can identify outsiders perfectly well.

As current field of research is quite unexplored, therefore not much work is done to analyze ties in online social networks. Hence, we analyzed the applicability of mixed data measures on small samples of Epinions network dataset and found that PN measure out of all has performed quite well in identifying outsiders in online networks also. Although a number of methods are available for analyzing ties in negative as well as positive tie networks but still there are some areas that need to be worked upon:

  i. Most of the measures discussed in paper, are analyzed by applying on standard datasets containing old and offline data. These measures need to be studied in recent data of online social networks such as Facebook, Twitter and LinkedIn.
 ii. PN measure can be further improved by varying value of β (attenuation factor) for better analysis of the behavior of relations in online social networks.
iii. Measures for analyzing ties based on Freeman's degree and betweenness views of centrality need to be further explored for its application in negative ties networks by using new approaches, such as Graph complement.
 iv. A combination of eigenvector centrality and Freeman centrality measures can be made to develop new measures that will exploit properties of both kinds of indices.
  v. The concept of negative clique can be further explored for developing new measures based on group behavior in negative ties network.
 vi. More optimized results can be achieved by first identifying cliques in social networks and then applying measures such as PN centrality for identifying actors showing negative behavior.

References

[1] Schneider F, Feldmann A, Krishnamurthy B, Willinger W. Understanding online social network usage from a network perspective. In: Proceedings of the 9th ACM SIGCOMM conference on internet measurement conference; 2009. p. 35–48.

[2] Kumar MJ. Expanding the boundaries of your research using social media: stand-up and be counted. IETE Tech Rev 2014;31 ():255–7.

[3] Davis JA. Clustering and structural balance in graphs. Hum Relat 1967.

[4] Heider F. Attitudes and cognitive organization. J Psychol 1946;21 ():107–12.

[5] Cartwright D, Harary F. Structural balance: a generalization of Heider's theory. Psychol Rev 1956;63(5):277.

[6] Doreian A, Mvar P. Partitioning signed networks. Soc Networks 2009;31:1–11.

[7] Boyd JP. Social semi-groups: a unified theory of scaling and blockmodeling as applied to social networks. Fairfax, VA: George Mason University Press; 1990.

[8] Borgatti SP, Daniel JB, Halgin DS. Social network research: confusions, criticisms, and controversies. Res Sociol Org 2014;40:1–29.

[9] Bohn A, Buchta C, Hornik K, Mair P. Making friends and communicating on Facebook: implications for the access to social capital. Soc Networks 2014;37:29–41.

[10] Box-Steffensmeier JM, Christenson DP. The evolution and formation of amicus curiae networks. Soc Networks 2014;36: 82–96.

[11] Smith JA, McPherson M, Lovin LS. Social distance in the United States Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004. Am Sociol Rev 2014;79(3):432–56.

[12] de Jong JP, Curşeu PL, Th AJ Leenders R. When do bad apples not spoil the barrel? Negative relationships in teams, team performance, and buffering mechanisms. J Appl Psychol 2014;99():514.

[13] Lakkaraju K. Reducing diffusion time in attitude diffusion models through agenda setting. In: Proceedings of the 2015 international conference on autonomous agents and multiagent systems, international foundation for autonomous agents and multiagent systems; 2015.

[14] Zsófia B, Néray B. Inter-ethnic friendship and negative ties in secondary school. Soc Networks 2015;43:57–72.

[15] Roethlisberger FJ, Dickson WJ. Management and the worker. Cambridge: Cambridge University Press; 2003.

[16] Read K. Cultures of the central highlands, New Guinea. Southwestern J Anthropol 1954:1–43.

[17] Sampson S. Crisis in a cloister. Unpublished PhD Thesis. Cornell University; 1969.

[18] Burke M, Kraut R. Mopping up: modeling wikipedia promotion decisions. In: Proc CSCW; 2008.

[19] Guha RV, Kumar R, Raghavan P, Tomkins A. Propagation of trust and distrust. In: Proc 13th WWW; 2004.

[20] Massa P, Avesani P. Controversial users demand local trust metrics: an experimental study on epinions.com community. AAAI '05. AAAI Press; 2005. p. 121–6.

[21] Brzozowski MJ, Hogg T, Szabó G. Friends and foes: ideological social networking. In Proc 26th CHI; 2008.

[22] Kunegis J, Lommatzsch A, Bauckhage C. The slashdot zoo: mining a social network with negative edges. In: Proc 18th WWW; 2009. p. 741–50.

[23] Lampe C, Johnston E, Resnick P. Follow the reader: filtering comments on slashdot. In: Proc 25th CHI; 2007.

[24] Breiger RL, Boorman SA, Arabie P. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. J Math Psychol 1975;12:328–83.

[25] White DR, Reitz KP. Graph and semigroup homomorphisms on networks of relations. Soc Networks 1983;5(2):193–234.

[26] Everett MG, Borgatti SP. Regular equivalence: general theory. J Math Sociol 1994;19(1):29–52.

[27] Labianca G. Negative ties in organizational networks. Res Sociol Org 2014;40:239–59.

[28] Lorrain F, White HC. Structural equivalence of individuals in social networks. J Math Sociol 1971;1(1):49–80.

[29] Hubert L, Schultz J. Quadratic assignment as a general data analysis strategy. Br J Math Stat Psychol 1976;29(2):190–241.

[30] Dekker D, Krackhardt D, Snijders TAB. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. Psychometrika 2007;72(4):563–81.

[31] Everett MG, Borgatti SP. Networks containing negative ties. Soc Networks 2014;38:111–20.

[32] Bavelas A. A mathematical model for group structures. Hum Org 1948;7(3).

[33] Leavitt HJ. Some effects of certain communication patterns on group performance. Unpublished PhD thesis. Massachusetts Institute of Technology, Cambridge; 1949.

[34] Smith SL. Communication pattern and the adaptability of task-oriented groups: an experimental study. Cambridge, MA: Massachusetts Institute of Technology; 1950.

[35] Bavelas A. Communication patterns in task-oriented groups. J Acoust Soc Am 1950.

[36] Cohn BS, Marriott M. Networks and centres of integration in Indian civilization. J Soc Res 1958;1(1):1–9.

[37] Pitts FR. A graph theoretic approach to historical geography. Professional Geogr 1965;17(5):15–20.

[38] Beauchamp MA. An improved index of centrality. Behavioral Science 1965;10(2):161–3.

[39] Mackenzie KD. The information theoretic entropy function as a total expected participation index for communication network experiments. Psychometrika 1966;31(2):249–54.

[40] Rogers D. Sociometric analysis of interorganizational relations: application of theory and measurement. Rural Sociol 1974.

[41] Freeman LC. Centrality in social networks conceptual clarification. Soc Networks 1979;1(3):215–39.

[42] Shaw ME. Group structure and the behavior of individuals in small groups. J Psychol 1954;38(1):139–49.

[43] Garrison W. Connectivity of the interstate highway system. Pap Region Sci 1960;6(1):121–37.

[44] Faucheux C, Moscovici S. Etudes sur la creativite des groupes taches, structures des communications et reussite. Bull C.E.R.P. 1960;9:11–22.

[45] Mackenzie KD. Structural centrality in communications networks. Psychometrika 1966;31(1):17–25.

[46] Czepiel JA. Word-of-mouth processes in the diffusion of a major technological innovation. J Market Res 1974:172–80.

[47] Nieminen UJ. On the centrality in a directed graph. Soc Sci Res 1973;2(4):371–8.

[48] Nieminen J. On centrality in a graph. Scandinavian J Psychol 1974;15:322–36.

[49] Kajitani T, Maruyama Y. Functional expression of centrality in a graph – an application to the assessment of communication networks. Electron Commun Jpn 1976;5924:9–17.

[50] Borgatti SP. Centrality and AIDS. Connections 1995;18(1):112–4.

[51] Borgatti SP. Centrality and network flow. Soc Networks 2005;27():55–71.

[52] Shimbel A. Structural parameters of communication networks. Bull Math Biophys 1953;15(4).

[53] Freeman LC, Borgatti SP, White DR. Centrality in valued graphs: a measure of betweenness based on network flow. Soc Networks 1991;13(2):141–54.

[54] Stephenson K, Zelen M. Rethinking centrality: methods and examples. Soc Networks 1989;11(1):1–37.

[55] Ford LR, Fulkerson DR. A simple algorithm for finding maximal network flows and an application to the Hitchcock problem. Rand Corporation; 1955.

[56] Ford LR, Fulkerson DR. Maximal flow through a network. Can J Math 1956;8(3):399–404.

[57] Ford LR, Fulkerson DR. Flows in networks. Princeton; 1962.

[58] Sabidussi G. The centrality index of a graph. Psychometrika 1966;31(4):581–603.

[59] Moxley RL, Moxley NF. Determining point centrality in uncontrived social networks. Sociometry 1974:122–30.

[60] Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol 1972;2(1):113–120.

[61] Ruhnau B. Eigenvector-centrality—a node-centrality? Soc Networks 2000;22(4):357–65.

[62] Bonacich P. Some unique properties of eigenvector centrality. Soc Networks 2007;29(4):555–64.

[63] Bonacich P, Lloyd P. Calculating status with negative relations. Soc Networks 2004;26(4):331–8.

[64] Bonacich P. Power and centrality: a family of measures. Am J Sociol 1987:1170–82.

[65] Katz L. A new status index derived from sociometric analysis. Psychometrika 1953;18(1):39–43.

[66] Hubbell CH. An input–output approach to clique identification. Sociometry 1965:377–99.

[67] Forsyth E, Katz L. A matrix approach to the analysis of sociometric data: preliminary report. Sociometry 1946:340–7.

[68] Festiger L. The analysis of sociograms using matrix algebra. Hum Relat 1949.

[69] Cook KS, Emerson RM, Gillmore MR. The distribution of power in exchange networks: theory and experimental results. Am J Sociol 1983:275–305.

[70] Leavitt HJ. Some effects of certain communication patterns on group performance. American Psychological Association; 1951.

[71] Berkowitz L. Personality and group position. Sociometry 1956: 210–22.

[72] Shaw ME. Communication networks. Adv Exp Soc Psychol 1964;1:111–47.

[73] Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Soc Networks 2001;23(3):191–201.

[74] Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Soc Networks 2006;28(4):466–84.

[75] Luce RD, Perry AD. A method of matrix analysis of group structure. Psychometrika 1949;14(2):95–116.

[76] Everett MG, Borgatti SP. Analyzing clique overlap. Connections 1998;21(1):49–61.

[77] Borgatti SP, Everett MG, Freeman LC. UCINET 6 for Windows: software for social network analysis (Version 6.102). Harvard, MA: Analytic Technologies; 2002.

[78] Borgatti SP. Identifying sets of key players in a social network. Comput Math Org Theory 2006;12(1):21–34.

[79] Smith JM, Halgin DS, Kidwell-Lopez V, Labianca G, Brass DJ, Borgatti SP. Power in politically charged networks. Soc Networks 2014;36:162–76.

[80] Emerson RM. Power-dependence relations. Am Sociol Rev 1962: 31–41.

[81] Emerson R. Exchange theory. Part II: exchange relations and networks. Sociol Theor Prog 1972;2:58–87.

[82] Markovsky B, Willer D, Patton T. Power relations in exchange networks. Am Sociol Rev 1988:220–36.

[83] Markovsky B, Skvoretz J, Willer D, Lovaglia MJ, Erger J. The seeds of weak power: an extension of network exchange theory. Am Sociol Rev 1993:197–209.