Journal of Computational Science 23 (2017) 91–108

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

# Ranking based comparative analysis of graph centrality measures to detect negative nodes in online social networks

Mankirat Kaur*, Sarbjeet Singh

*Computer Science Engineering, University Institute of Engineering and Technology, Panjab University, Chandigarh, India*

ARTICLE INFO

*Article history:*
Received 23 November 2016
Received in revised form 1 September 2017
Accepted 17 October 2017
Available online 31 October 2017

*Keywords:*
Online Social Networks (OSNs)
Centrality
Negative ties
Social network analysis

ABSTRACT

Online Social Networks (OSNs) are emerging as a communication platform where interaction among users results in the formation of positive or negative relations. Due to the existence of negative relations, many nodes are suspicious of masquerading and conspiring against popular nodes as well as intruding into the private groups of networks. Many researchers have analyzed negative nodes in networks of positive and negative ties by using measures such as degree, status, PII and PN centrality. While the existing literature focused only on small offline datasets, in this work an approach to identify negative nodes in large datasets of OSNs is proposed. The deviation of results of measures from actual behavior is examined using statistical and graphical techniques. It was observed that PN centrality measure is able to detect a number of outsiders of the network with higher accuracy as compared to other measures. However, some crucial nodes which are actually outsiders are misclassified as the most popular nodes by it. To counter this drawback, we have proposed and compared new values of parameters of PN measure for large-scale networks through graphs as well as statistical measures.



## 1. Introduction

OSNs are drawing the attention of the large community, where each user can create their personalized profile which usually consists of their identifying information, interest groups, and personal contacts. The main functionality of OSNs is to bring people with diverse beliefs and notions at the common interface by making them interact with each other through messaging service. Communication patterns among users in the social network often result in developing positive and negative relations. Positive relations are formed by friendship, endorsement, liking, and trust, whereas, negative relations are a result of disliking, distrust, opposition, and antagonism. The identification of negatively behaving malicious nodes in social networks, who are responsible for breaching of secret information by intruding into private networks through analyzing the structure and connections of nodes, has become an eminent area of research. Negative ties are a part of various network analysis approaches such as balance theory [1,2], clusterability theory [3], their generalizations, blockmodelling analysis [4], and semigroup work [5]. Formerly, when the network among nodes was societal and small, the ties were analyzed by using the standard concepts of network analysis such as equivalence of nodes, centrality and influence measures. The most influential and central node through which most of the network traffic flows were identified through some basic calculations on their immediate connections. Many standard datasets also described the negative relations in these societal networks such as skirmish relation in “bank wiring room” data, reported by Roethlisberger and Dickson [6]; enmity “hina” relation, reported by Read in 1954 [7] in social network of tribes of the Gahuku–Gama alliance structure of the Eastern Central Highlands of New Guinea; the disesteem and disliking relations among group of monks reported by Sampson in 1969 [8].

As the technology evolved, the social networks are switched to online web based systems. The malicious and illegal activities of spams and crimes, such as intrusion into computer systems, masquerading, cyber bullying, cyber terrorism, etc. which are somehow related to negative relations among users showed a sharp increase [14–16]. Moreover in today’s scenario, where social interactions are carried out online and profile of an intended person can be fake with whom one wants to communicate, therefore, exploration of such type of relations in social networks has become the need of the hour. Such types of relations can be found in websites, such as positive and negative comments posted by users about the product being reviewed in online rating websites like Epinions.com [9,10] and tagging other users as friend or foe in a comment posted by them in Slashdot.com [11–13].

* Corresponding author.
*E-mail address:* mannsunshine09@gmail.com (M. Kaur).

https://doi.org/10.1016/j.jocs.2017.10.018
1877-7503/

The negative nodes are also suspicious as email spammers in email communication network and their activity pattern needs to be monitored. Such negatively behaving nodes are also considered as anomalous nodes. This analysis shows strong correspondence to anomaly detection in social networks [17–20].

There are many differences in characteristics of social processes that occur in positive and negative ties networks, such as:

i Transitivity of flow: Positive relations form a network of well-connected nodes through which information is quickly delivered to a group of nodes via the direct and indirect paths, thus depicting a high level of transitivity. Whereas in a network containing negative relations, the information diffuse hardly from one node to another attributing to the low level of transitivity.

ii Sparseness: The graphs of positive networks are very dense but negative networks form disconnected graphs due to which centrality approaches (used to find most central or influential node of the network) becomes difficult to apply.

For analyzing ties in social networks, researchers have developed measures which include Freeman's degree, betweenness and closeness centrality measures [21], Katz influence measure [22], Hubbel measure [23], Bonacich power measure [24], etc. But all these measures are only suitable for analyzing positive relations as they all depend on the flow of information among nodes of the network, which is absent in the case of negative ties.

Due to the difference in flow characteristics of negative ties, many new measures are designed that are applicable to both positive and negative tie networks. These are Degree, status [25], PII [26] and PN centrality [27]. Each measure is based on the different concept to identify the most influential and most negative node of the network. The effectiveness and accuracy of each measure for analyzing negative nodes were earlier examined by applying them on small datasets, such as Sampson's monastery [8] and Bank Wiring Room [6] by Everett and Borgatti [27]. These datasets were gathered by original researchers physically present at locations, inferring the relationship between nodes by observing their behaviors. Such datasets contain structure and connection behavior of thirty nodes at maximum. The analysis made by Everett and Borgatti [27] concluded that out of four mixed data measures, PN centrality is the only measure that can uniquely identify negative nodes in small datasets captured by researchers. No such assessment was made for large-scale online networks. The results need to be investigated further by applying them to large datasets of current online social networks with thousands of nodes interacting with each other. Also in such larger networks, the results obtained from each measure can be analyzed to figure out which measure is giving better results in which type of online network.

The following are the main contributions of this paper:

1. Unlike offline datasets, it is very difficult and challenging in online networks to collect the dataset with predefined labelling. To overcome this difficulty, an approach is proposed to collect and label data for analyzing negative ties in online networks. The accuracy of each measure is verified by comparing the ranking of nodes given by each index with their Actual ranks by using graphs in three datasets of Epinions, Slashdot and Wikipedia websites [9–13].
2. The statistical measures such as Kendall rank correlation coefficient, Chi-square test for goodness of fit and F-score are also derived for determining the association between ranking provided by measure with their Actual Ranks as well as to analyze how well the measure actually reflects the expected behavior. This helps us to identify which measure (out of the four) is appropriately ranking the centrality of nodes and is able to identify most of the outsiders.
3. The ranking given by PN centrality closely matches with the Actual Ranking as compared to other measures but shows slight variation. In defiance of this shortcoming, we have also proposed new values of the parameter $\beta$ (normalization factor) of PN measure that can withstand large scale datasets and can produce the most desired results.

The paper is organized as follows. Section 2 discusses the related work which includes standard methods of analyzing ties, their limitations in negative networks and measures developed specifically for analyzing both positive and negative ties. Section 3 consists of an overview of limitations of existing study and the need for further investigation. Section 4 introduces an approach for analyzing negative ties in larger datasets of online networks. The description of datasets used and experimental results obtained by using proposed approach are given in Section 5. Finally, the paper is concluded with future directions in the current field.

## 2. Related work

The analysis of ties in social networks has been a prominent area of research for a long time. Examining the social processes that may occur in positive and negative ties of the network helps in exploring different aspects of social networks. The vast literature prevails in the field of identifying ties among actors of the social network. Many methods were proposed to analyze the relations existing among persons and to calculate the popularity of each individual. As already discussed in Section 1, the social processes that may occur in positive ties may or may not take place in negative ties networks due to differences in flow and density of their networks. Many standard network concepts are directly applicable to negative ties network without any modification. These concepts include role or positional equivalence such as structural equivalence [28–30], regular equivalence [29,30] and automorphic equivalence [28,30]. These also include statistical techniques like QAP correlation [31] and regression [32]. The equivalence concepts are used to find nodes that are having similar patterns of relations with other nodes of the network and thus playing same roles and positions. Such concepts are equally applicable to both negative and positive relations.

The statistical techniques are used to check the similarity or correlation among various parameters. These can be used in both positive and negative ties. For example, in a positive relation such as friendship with colleagues within an organization can be positively correlated with an aspect of the sameness of states to which they belong. It can also be correlated with the factor of promotion on the job as friendship with colleagues may help in providing critical information and references. Similarly, in the negative relation of disliking colleagues at a job can be positively correlated with the demotion of an employee in organization structure as colleagues may provide negative feedback about a person to a higher authority. The only difference between the required models of correlation in both ties is promotion and demotion of an employee, respectively. Rest all factors like the sameness of language, state or region, age etc. can be modeled in the same manner for both negative and positive relations.

Three types of centrality measures, namely Degree, Betweenness, and Closeness were developed by Freeman, which was a bit more complex. Degree centrality is the measure of influence to calculate the immediate risk of getting infected from directly connected nodes in one time period in an infected network of nodes [33]. In positive tie network, degree measure is used to calculate the popularity of node, i.e. the node which is receiving more number of positive ties from other nodes in the network (high in-degree) and sending many ties to other nodes(high out-degree) is the most adored person in the network. A similar interpretation can also be made in negative ties network of disliking relationships. The most disliked individual in the network is the one 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.

The measures of betweenness centrality have focused on the frequency of arrival of packets. It 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 be present at two places at one time, that is through duplication. Second, traffic can take only shortest paths and flows have a predefined origin and target [34]. Betweenness measures can be used in positive ties where a high level of transitivity and flow of information exists. But as discussed, the flow of resources is absent in negative tie networks, therefore, it becomes very difficult to calculate betweenness measures in such type of networks.

Closeness measure is the interpretation of expected arrival time of something flowing through the network only if it follows the shortest path or follows parallel duplication where things flow parallel through all possible paths [34]. Also, closeness can be computed only in connected graphs with the flow-based network. Therefore, this measure can easily be applied on positive ties where transitivity is maximum, like friends of friends are friends. But in the case of negative ties networks, which are usually represented by disconnected graphs, closeness becomes zero as nodes are not reachable from one another. Also, things do not flow for more than one length path in negative ties, so closeness is difficult to be applied for computing node centrality.

Apart from Freeman's three centrality measures [21], Katz influence [22], Hubbel measure [23], and Bonacich power measure [24] were also proposed to analyze the impact of the popularity of neighboring actors on the centrality of given node.

Most of the above-mentioned measures are only appropriate for positive networks. They cannot be applied to analyze negative ties as they depend upon transitivity of flow of information in the well-connected network which is very less in the case of negative ties. Therefore, a number of new measures that comply with the structure and conditions of negative ties networks were developed which are described below:

1. Degree measure: This measure calculates the degree of an actor as the total number of connections that an actor have with other actors of the network. The value of the degree of actor *i* can be computed by adding all the values in the $i^{th}$ row of adjacency matrix that corresponds to the connections of actor *i*.
2. Status measure: Bonacich and Lloyd [25] proposed status measure using eigenvectors. According to this measure, the status of an actor in a network of positive and negative ties rises if it has positive connections with most of the popular actors. Also, its status may also decrease by having negative connections with the same. If an actor has positive relations with low-status actors in the network then also its status score decreases.
3. PII (Political Independence Index): Smith et al. [26] proposed this new measure for analyzing the power of an actor in political networks of positive and negative relations. According to this measure, any actor (focal actor) which is dependent on other actors (also called adversaries) with respect to control of the flow of resources, information and support can increase the power of adversaries by losing its resources to them. When adversaries of focal actor grow then it makes alliances with other weak actors which are also in direct threat and thus reduce its dependency on adversaries.
4. PN centrality: Everett and Borgatti [27] proposed this measure for identifying ties in a mixed network of positive and negative relations. It is based on the concept that having negative connections with a most disliked person is better than having negative ties to a most central person of the network. For normalizing positive and negative ties, β (attenuation factor of this measure) was chosen as *1/(2n-2)* in the complete network where *n* is a total number of nodes and matrix A is calculated as A = P-2N, P is positive tie matrix and N is negative tie matrix. PN is expressed as:

$$PN = \left(I - \beta A\right)^{-1} 1 = \left(I - \frac{1}{2n-2} A\right)^{-1} 1 \tag{1}$$

Everett and Borgatti [27] analyzed the accuracy of these measures in identifying outsiders in social networks by applying them on a dataset of Sampson's monastery [8]. They concluded that out of these four measures, PN centrality is able to uniquely identify and rank the outsiders in a social network, whereas degree, status and PII measures are able to identify most of the negative nodes but cannot provide the unique score to every node of the network. Other types of studies have also been carried out to calculate centrality in online social networks based on trust relationships [35–37] but, in our work, we are mainly focusing on determining negative relations in online networks.

## 3. Problem overview

The conclusions given by Everett and Borgatti [27] from their experimental study on negative ties in small datasets are:

1. Both the PN and Degree measures are able to identify outsiders of network efficiently by providing low scores to them.
2. The degree is not a very sensitive measure as it provides the same score to many actors but same is not true for PN which gives every actor a unique score.
3. Status measure fails to identify outsiders if more than two groups of actors exist as it assumes that any data can be partitioned into only two groups.
4. PII measure is not able to capture the true sense of centrality as it only analyzes the power that each actor possesses in presence of allies and adversaries.

Following limitations are observed in work carried out by Everett and Borgatti [27]:

1. The study made by them is limited to offline datasets that contain only 20 nodes.
2. The accuracy of results obtained from all measures is not validated on different online datasets.
3. The change in interpretation of results of measures in case of different types of datasets is not analyzed.

The work presented in this paper is carried out according to the above mentioned limitations.

## 4. Proposed approach

To achieve the contributions of the paper discussed in introduction and to mitigate limitations stated in Section 3, an approach is proposed to analyze negative ties in large online networks. The effectiveness of each measure in identifying negative nodes in large networks and the changes in the behavior of measure on the transition from small to large datasets are analyzed. Further, the results of

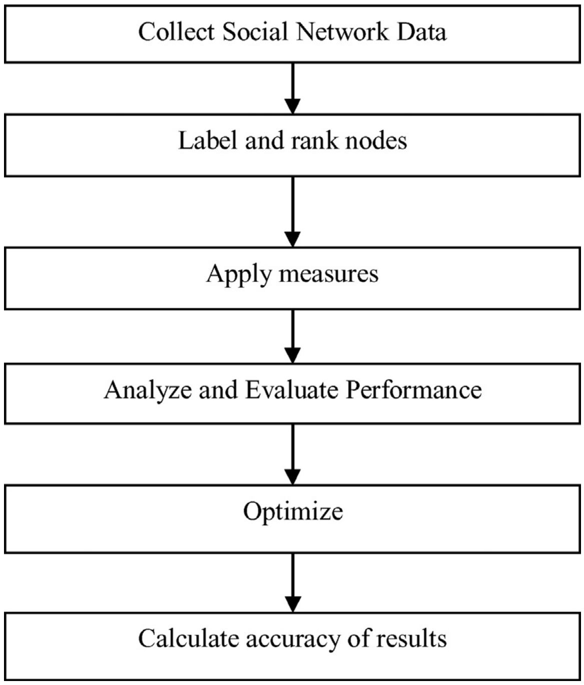


**Fig. 1.** Overview of Proposed approach.

large-scale datasets are optimized by varying the values of parameters of the best index. This helps us to find which measure is appropriately identifying outsiders in larger networks and is easily adaptable to them.

For analyzing negative ties in large networks, online datasets containing positive and negative ties among nodes of the network are collected. The effectiveness of each measure is analyzed by using graphs. The graph of the rank of nodes versus node numbers is plotted. From such a large dataset, a definite set of nodes is extracted whose ranks are assigned inaccurately by most of the measures (i.e. the ranking of these nodes shows deviations from Actual Ranking for each measure). A graph of ranking provided by each measure for these set of nodes is plotted with Actual Ranking and the deviation from latter is detected. The curve of measure which lies very close to Actual Ranking is considered as the most appropriate measure for identifying negative ties.

In Social Network Analysis, a network is depicted as a graph G(V, E) where V is the set of vertices representing nodes or actors and E is the set of edges or connections between a pair of nodes in the social network. The connections between a pair of nodes are depicted in the form of adjacency matrix A, in which element $A_{ij}$ represent the weight/value of connection between node i and node j. The ties are represented by the value of element $A_{ij}$, as greater than zero for the positive tie, less than zero for the negative tie and zero for no connection. This adjacency matrix A is also called signed matrix and graph containing positive and negative values of edges is called signed graph. Sometimes, the signed graph is also represented in two different matrices. The information about positive connections is stored in a positive matrix (P) and negative connections are stored in a negative matrix (N).

The proposed work is accomplished by going through steps described in flow chart of Fig. 1.

i Social Network Data Collection

The large dataset of online social networks is gathered consisting of positive and negative edges among nodes, also known as signed network data. Such dataset contains a substantial number of nodes interacting with each other. In order to simplify the analysis of different measures on online data, a sample of few thousand nodes is extracted from the large dataset using systematic sampling technique. The data about the connection between nodes is collected in three types of matrices, namely, A, P, and N to satisfy input requirements of different measures. In this paper, datasets of Epinions [9,10], Slashdot [11–13] and Wikipedia [38,39] websites are used for analysis purpose.

ii Labeling and Ordering of nodes

Most of the datasets of online networks available for analysis purpose are not labeled to indicate the most negative and positive nodes of the network. In order to verify the accuracy of measures, a predefined ranking of nodes is needed with which ranking given by all measures can be compared. This ranking is achieved by analyzing the connections of a given node with other nodes of the network. The basic concept of centrality is used to analyze the rank of each node in the network and then all nodes are placed in increasing order of their centrality.

For example, a node which is having positive ties with highly negative (disliked) nodes attains low centrality as compared to one which has positive ties with highly positive (popular) nodes of the network. Similarly, if a node has negative ties with highly negative nodes, it gains high centrality as compared to the one which has negative ties with highly positive nodes of the network. In other words, the node having a maximum number of negative connections with popular nodes of the network is assigned the first rank and the node with maximum positive ties with other nodes (i.e. the most positive node) is given the last rank. This ranking is then named as Actual Ranking which is used for comparing the accuracy of ranks given by other measures. Since this task is very difficult and complex in the case of datasets containing thousands of nodes, therefore, a group of experts in the area of Social Network Analysis is selected to rank the nodes. Accordingly, Actual Ranking is determined through a collective decision, which is a result of similarity in ranking orders given by the majority of experts.

iii Applying measures

The sample datasets gathered in step 1 are imported in social network analysis software UCINET. The Degree, Status, PII and PN measures are applied on datasets for identifying the negative nodes. The resultant set of each measure is then obtained by correctly specifying the parameter values that were used by Everett and Borgatti [27] for analyzing results in small datasets. By doing this, it can be verified whether the results obtained from large datasets remain same as small datasets or not, and the differences between both results are identified. The four measures are described as follows.

a a. Degree measure: It is calculated by subtracting a total number of negative ties of the node from its total positive ties:

$$D(m) = P(m)\text{-}\mathrm{N}(m) \tag{2}$$

where $P(m)$ is a number of direct positive ties, $N(m)$ is a number of direct negative ties and $D(m)$ is a degree of node $m$.

b Status measure: The status score of each node correspond to each element of eigenvector. It is calculated as follows:

$$Ay = \lambda y \tag{3}$$

where $A$ is the adjacency matrix of the network, $\lambda$ is the largest eigenvalue and $y$ is the eigenvector.

c Political Independence Index: It calculates the power of node in a political network of allies and adversaries. It is described as follows [26]:

$$\sum_{i=0}^{k} \beta^i \left[ P(i)^x - N(i)^x \right] \tag{4}$$

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

$$x \le \frac{\ln(2) - \ln\left(|\beta|\right)}{\ln(M)} \text{i.e. } \beta M^x \le 2 \tag{5}$$

Here, $M$ is the maximum number of ties from any node in the network.

d PN Centrality: The fourth measure is based on calculating centrality in mixed networks of positive and negative ties. It is calculated as [27]:

$$PN = \left[ I - \frac{1}{2n-2} A \right]^{-1} 1 \tag{6}$$

Here $A$ = P-2*N, β = 1/(2n-2) is the attenuation factor to normalize the positive and negative ties of the network.

iv Analysis and Evaluation

In this step, the different results obtained from step 3 of all measures are compared with Actual Ranking of nodes derived from step2. The effectiveness of each measure is analyzed with the help of the graphs. A graph serves as a better method for visually analyzing and comparing the ranking of measures with respect to Actual Ranking. The number of nodes in a sample is very large due to which the graph turns out to be congested and analysis becomes difficult. Therefore, for better visualization, a subset of nodes is selected from the sample for which ranking given by most of the measures shows deviation from the Actual Ranking. A graph between Ranks of nodes and Node numbers is plotted. The ranks are allocated to nodes of the network in increasing order of scores given by each measure. This means the first rank is given to a node which is having lowest score i.e. the most negative node of the network and the last rank is given to the most positive node having the highest score. The numbers to nodes are assigned in such a way that Actual Ranking forms a straight line. It means that a node with the first rank is assigned with node number 1, the second rank is assigned node number 2 and so on. The graph of ranking of a subset of nodes is plotted for all resultant sets along with Actual Ranking. The curve of measure which is lying very near to actual curve is then considered as an effective measure for analyzing ties. The behavior of measures which are showing fluctuations from Actual Ranking are observed and the factors responsible for such deviations are also discussed.

v Optimization

The selected efficient measure from step 4 is then explored by varying values of parameters. Those values of parameters are preferred that are compatible with the properties of larger networks and are able to remove fluctuations that are observed in the results of previous parameter values. The result sets of different values of parameters are plotted along with Actual Ranking curve. The value of the parameter for which ranking of nodes nearly matches with Actual Ranking is considered as the most efficient value of the desired measure for identifying negative nodes in large datasets of online social networks.

vi Calculation of accuracy of Results

In step 4, the results are evaluated through graphs for visually analyzing the deviations from the actual curve. For this purpose, a subset of the sample is extracted and examined. However, for comparing the results from each measure with Actual one over whole sample size, the statistical techniques are required which will articulate the correlation among them accurately irrespective of the size of the data. Therefore, three types of statistical measures are used as described below:

a) Kendall Rank Correlation Coefficient: It is commonly referred to as Kendall Tau's Coefficient. It is used to check the degree of independence between two variables by evaluating the two sets of ranks provided to same objects by these two variables. This statistic can be calculated as:

$$\Gamma = \frac{(\text{no.ofconcordantpairs}) - (\text{no.ofdiscordantpairs})}{n(n-1)/2} \tag{7}$$

where, Concordant pairs = Number of ranks ordered in the same way

Discordant pairs = Number of ranks ordered differently

$n$ = sample size

If there is a perfect association between two rankings, then the value of Kendall coefficient correlation is 1; if two variables are independent, correlation is 0 and for perfect non-association, it is −1. It can be used to detect the correlation between the ranking of each measure and the actual one. The measure with the highest value of Kendall coefficient will be regarded as the best index for identifying negative nodes.

b) Chi-Square Test of goodness-of-fit: This measure is used to detect how well the observed data of categorical variable fits the expected model. Basically, it tests the association between two variables and is calculated as follows:

$$X^2 = \Sigma \frac{(observed - exep\,eccted)2}{expected} \tag{8}$$

where, observed values are ranking order of each measure and expected values are their corresponding actual ranking order. If the value of the $\chi^2$ statistic is less than its tabulated value at given a degree of freedom and particular significance level, then it can be concluded that the observed values of measure fits the actual values and measure with the lowest value of this statistic will be considered as the best fit to expected actual ranks of nodes.

c) F-score: In a statistical analysis of binary classification, F-score is a measure to determine the accuracy of the method by accounting both precisions and recall to compute the final score. In our study, the precision is calculated by dividing the number of correctly predicted negative nodes with the sum of a number of positive nodes predicted as negative and correctly predicted negative nodes. However, recall is the number of correctly predicted negative nodes divided by sum of a number of negative nodes wrongly predicted as the positive and total number of correctly predicted negative nodes. The precision and recall can be calculated in terms of true positives, false positives and false negatives as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{9}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{10}$$

The balanced F-score is the harmonic mean of precision and recall, where F-score reaches its best value at 100 and worst value at 0. It is used to compare the efficiency of different measures in identifying the outsiders of network and is computed as:

$$F_{score} = \frac{2 * \text{True Positives}}{2 * \text{True Positives} + \text{False Positives} + \text{False Negatives}} \quad (11)$$

These six steps of approach are repeated for multiple datasets and values of parameters that are able to optimize results in these datasets are identified. By selecting multiple datasets of online networks, the behavior of each measure in different networks can be studied and better analysis can be performed on the application of each measure.

## 5. Experimental results and comparison

### 5.1. Description of datasets

The two datasets are collected from SNAP (Stanford Large Network Dataset Collection) website for analyzing ties [40]. The signed network datasets are used for identifying negative relations. The datasets contain three types of values of edges between nodes viz. +1 value of edge depicts the positive relation, such as trust or friend, 0 value of edge depicts no relation and −1 value depicts negative relation between nodes, such as distrust or foe. The relations described in datasets are collected from online rating websites where users give their reviews about the product and other users show their trust or distrust in the reviews as well as on product reviewed. There are three types of online review website datasets collected for the experimental purpose.

#### 5.1.1. Epinions signed social network dataset

This is the online social trust network of who trusts whom, of a common consumer review website Epinions.com, where users create signed relations of trust or distrust with each other [9]. Members of the site can give positive or negative ratings to products being displayed as well as rate the reviews given by other members. The visitors of the site can check the new and old reviews of the product and then decide which product to purchase. The links between nodes or persons who give reviews about products are explicitly labeled as positive or negative. The positive value of the link between nodes $u$ and $v$ depicts that the review given by $v$ is trusted by $u$ and vice-versa. All the relationships formed by trust results in the formation of Web of Trust, which is then combined with review ratings to choose which review should be shown to the user [10]. The dataset of Epinions trust network is available on the website of SNAP. The +1 value of edge indicates the positive relation of trust and −1 value indicates the negative relation of distrust.

#### 5.1.2. Slashdot signed social network, November 2008 dataset

Slashdot is a technology-related news website that entails specific user community. The news related to current technology is submitted by users which editors of site evaluate and provide open discussion among readers. Each news article also has a section for comments attached to it, where users of the site post their comments and lead to open threaded discussion. The comments associated with each news article are then rated by editors of sites by using moderation system. In this system, random moderators are appointed which uses points given to each comment (out of five) for rating purposes. The moderators rate each comment as +1 or −1. In 2008, Slashdot launched a new feature known as Slashdot Zoo which allows users to tag other users as friends or foes. The dataset used in this work contains links to friends (+1) or foes (-1) between users of the website [11–13]. The dataset was collected in November 2008 and is now available on the website of SNAP.

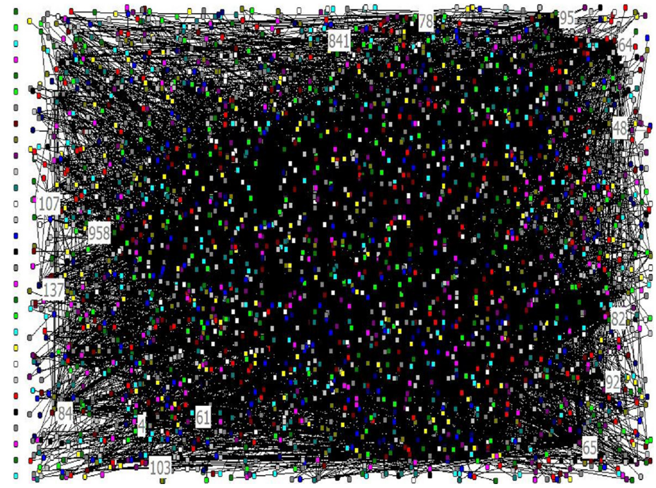

**Fig. 2.** The graph of dataset of Epinions signed social network.

#### 5.1.3. Wikipedia adminship election signed network dataset

Wikipedia is a free encyclopaedia, written by many of its readers, collaboratively, around the world. The website called wiki was designed to make the process of collaborating information easy. The special class of volunteers working on Wikipedia acts as administrators, who are provided with access to additional technical features of the website for maintenance purposes. A person who wants to become administrator needs to issue Request for Adminship and then the Wikipedia community will decide who will be promoted to adminship either through public discussion or through a voting mechanism. A dataset of administrator elections and vote history was extracted from Wikipedia page edit history from January 2008. This dataset constitutes 2800 elections with approximately 100,000 total votes and 7000 users who participated in the election process, either as a voter or as a contestant. Each edge in dataset consists of a positive (+1) or negative (-1) vote given to contestant by users of the website [38,39]. The collected dataset is now available on SNAP website [40].

### 5.2. Experimental results

The experiments to analyze ties are conducted on three datasets of OSNs by using proposed approach and results obtained are discussed as follows:

#### 5.2.1. Social network data collection

Generally, the size of datasets of OSNs is very large and exploring complete network for extracting features becomes a cumbersome job. This can be better analyzed from Figs. 2 and 3. In order to simplify the analysis of the behavior, connections, and structure of nodes, the size of datasets needs to be shortened by applying sampling techniques. The sample of 3606 nodes is extracted from Epinions dataset and that of 3616 nodes is extracted from Slashdot dataset using Random Walk algorithm of Scale Down sampling. As sample size is much smaller than actual size, Random Walk algorithm outperforms the Forest Fire and random PageRank algorithms [41]. However, the complete dataset of Wikipedia adminship election contains 7000 nodes which are of considerable size and can be analyzed directly without the need of sampling.

#### 5.2.2. Labelling and ordering of nodes

The next step involves labelling of nodes by an expert, as positive or negative of sample dataset by observing their structure and connections. For example, in Epinions network, a user whose reviews are trusted by most of the other users is considered as the most

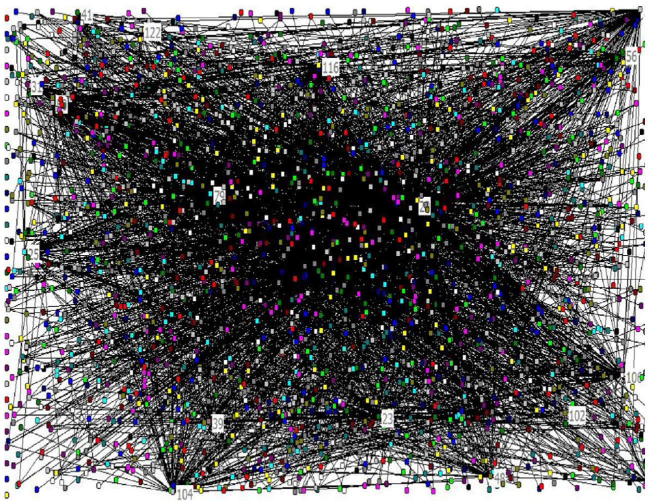

**Fig. 3.** The graph of dataset of Slashdot signed social network.

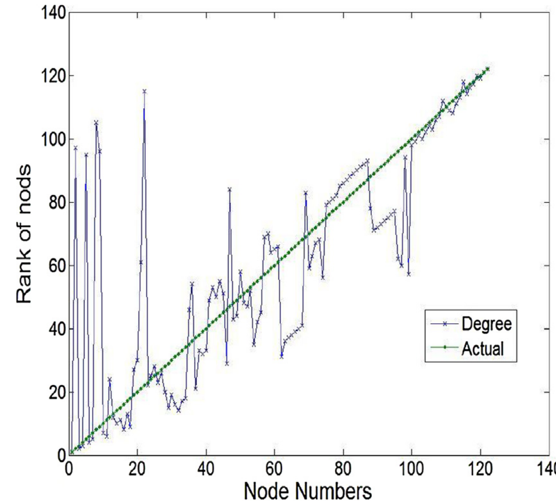


**Fig. 4.** Comparison of Degree Ranking versus Actual Ranking in Epinions signed social network.

trustful and authoritative user. However, a person whose reviews about products are distrusted by maximum users of website acts as the most distrustful or negative node of the network. In Slashdot network of the specific user community, the person who is tagged as a friend in comments on most of the news article discussions is considered as the most influential person and news articles submitted by him/her are the most reliable articles. Similarly, a person who is tagged as a foe by most of the readers of the website is considered as the most negative person and correspondingly articles submitted are considered as non-authenticated. In Wikipedia elections network, the person who bagged the maximum number of supportive (+1) votes from users is considered as the most popular node and promoted to adminship whereas the contestant with all opposing (-1) votes is the most disliked node in the network. The nodes are then ranked in increasing order of their centrality as shown in Tables A1–A3 respectively (See Appendix A). This ordering of nodes is then named as Actual Ranking which can be used for comparing the effectiveness of ranks given by measures.

### 5.2.3. *Applying measures*

This two sample and one complete dataset are then imported into UCINET software by converting them into DL file format. Then these DL files are converted by UCINET tool to already mentioned three matrices: signed adjacency matrix A, positive ties matrix P, negative ties matrix N. The four metrics named as Degree, Status, PII and PN centrality are computed. The results are presented in form of scores assigned to all nodes of the network by each measure.

### 5.2.4. *Analysis and evaluation*

The results obtained from different measures are examined and ranking is provided to all the nodes in increasing order of scores given by each measure. The ranking of nodes provided by all measures needs to be compared with Actual Ranking for checking their effectiveness in identifying negative nodes. For analyzing results visually in form of graphs, subsets of those 122 nodes are selected from Epinions and Wikipedia networks for which most of the measures provide different ranks as compared to the Actual one. A similar subset of 118 nodes is selected in case of Slashdot network. The rank number 1 is given to the most distrustful or disliked node of the network and the last rank i.e. 122 in Epinion and Wikipedia, and 118 in Slashdot, is given to a node which is the most trustful or popular. A graph is plotted with node numbers on x-axis and rank of nodes on the y-axis. The numbers to nodes are assigned in such a way that Actual Ranking forms a straight line on a graph. The node which is ranked at first position is the most negative node (i.e. N79 in Epinions, N2354 in Wikipedia and N46 in Slashdot, as shown in Tables A1–A3 in Appendix A) and is given a node number 1on x-axis. Similarly, node ranked at the second position is given number 2 and so on. The comparison of result sets of measures with Actual Ranking is summarized in following sections.

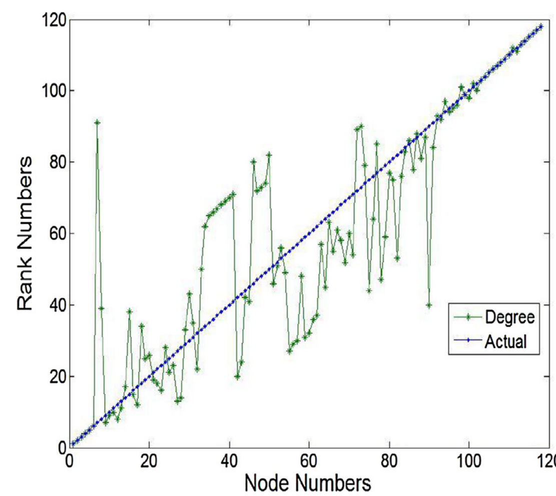


**Fig. 5.** Comparison of Degree Ranking versus Actual Ranking in Slashdot signed social network.

I) Degree measure vs. Actual Ranking

In Epinions signed network dataset, the results can be analyzed from Table A1 in Appendix A and Fig. 4. It is observed that ranks assigned by degree measure are very different from that of the Actual Ranking. Node N5 is ranked 97, that is, among the positive nodes of the network because it has positive degree 4. However, when we observe its structure, it has negative ties with 494 highly trusted nodes of the network, which according to centrality concept should be considered as extremely negative or distrusted node. Similarly, degree measure ranked many negative nodes of a network such as N52, N39, and N41 as positive nodes purely on the basis of positive degree score which does not provide a complete view of structure and centrality of a node. On a similar note, Slashdot signed network dataset can be analyzed from Table A2 in Appendix A and Fig. 5, and it is observed that ranks assigned by degree measure are very different from that of the Actual Rank-

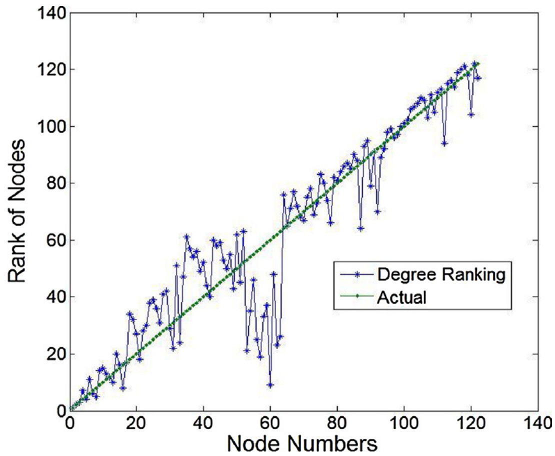


**Fig. 6.** Comparison of Degree Ranking versus Actual Ranking in Wikipedia signed social network.

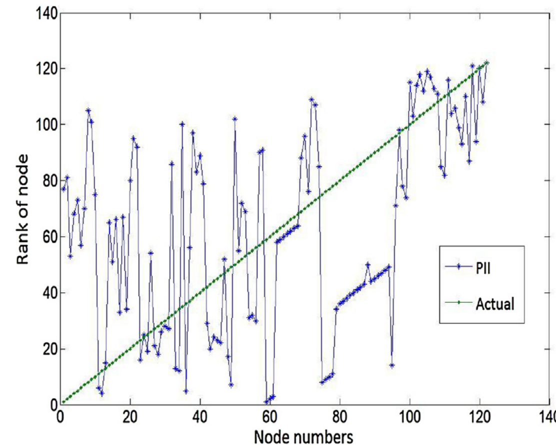


**Fig. 7.** Comparison of PII Ranking versus Actual Ranking in Epinions signed social network.

ing. Node N5 is ranked 97, that is, among the positive nodes of the network because it has positive degree 4. However, when we observe its structure, it has negative ties with 494 highly trusted nodes of the network, which according to centrality concept should be considered as extremely negative or distrusted node. Similarly, degree measure ranked many negative nodes of a network such as N52, N39, and N41 as positive nodes purely on the basis of positive degree score which does not provide a complete view of structure and centrality of a node. Likewise, the results for Wikipedia signed network as observed from Table A3 in Appendix A and Fig. 6 shows that N906 and N192 have the same degree of −18 but their order is shuffled in degree ranking. This is due to the fact that N906 has all 18 ties negative, whereas, N192 has 9 positive and 27 negative ties but degree measure fails to understand the internal connections of the node to deduce the actual rank of a node in the network. As we move from left to right along a degree graph the fluctuations reduce and it tries to overlap with the actual curve.

From these results, it is clear that behavior of degree measure for identifying negative nodes of the large network is completely different from that found in small datasets and the degree measure fails to capture the aspect of centrality that a node possesses through interactions with other nodes of the network.

II Status Ranking vs Actual Ranking

Status measure computes the score of each actor in the network by finding the eigenvalues of signed adjacency matrix A. Each element of the eigenvector corresponding to largest eigenvalue depicts the score of each actor of the network. Bonacich and Lloyd [25] stated that the values of the eigenvector of balanced symmetric matrix A of network form two cliques. The members of the first clique have positive scores whereas members belonging to the second clique have negative scores. All the positive ties exist among the members of the same clique and negative relations exist among members of other cliques. The actors having lower scores are the negative nodes of the network. But when eigenvalues of selected datasets are computed, it is observed that eigenvalues of signed adjacency matrix A are imaginary numbers and their all eigenvectors get zero value.

In such cases of very large datasets, the eigenvalues of the adjacency matrix of the network may not be real or there may exist multiple eigenvalues with the same magnitude. In such condition, it becomes very difficult to choose one eigenvalue and its corresponding eigenvector [27]. Even if we choose one eigenvalue, there may exist multiple eigenvectors corresponding to that eigenvalue. Out of them, it is very difficult to choose one eigenvector for calculating status scores. By using status score, one can identify negative nodes of small networks but in larger networks status measure may or may not identify them. In large scale networks, multiple groups of nodes may exist that are having intragroup positive ties and intergroup negative ties which cannot be identified by status measure as it divides the network into two cliques. As a result, its calculation is not time efficient. Therefore, the status measure is not suitable for analyzing ties in larger datasets.

III PII Ranking versus Actual Ranking

This measure actually determines the power of node in a mixed network of positive and negative ties. In Epinions signed social network dataset, it can be analyzed by observing the ranks given by PII measure in Table A1 in Appendix A and Fig. 7 that the most negative nodes of the network are ranked among the positive ones. It identifies node N42175 as the most negative node since it has a negative connection with the most powerful node N20 of the network (which is available with a large number of alternatives with which it can make trust relationships to increase its power). Further, N42175 has no connection with any other node of a network which makes it fully dependent on N20 and thus becomes least powerful by losing all its resources to this node. All the nodes which are having a similar type of connection with N20 are ranked among highly negative nodes of the network. Rank 4 is given to node N37772 because it is having distrust ties with three well-connected nodes of the network which makes it fully dependent on them and becomes powerless. Working on a similar note, in Slashdot signed social network dataset, PII measure is able to identify first ten outsiders of the network but after that, it starts ranking nodes very high, as shown in Table A2 in Appendix A. In the last part of the table, PII is ranking some of the positive nodes of the network among the highly negative nodes. It can be observed from Fig. 8 that PII is able to find negative nodes but fails to identify actor who is receiving most of the positive comments from readers on news article posted by him/her. The deviations of PII from Actual Ranking can be analyzed from the graph in Fig. 8. In a similar way, In Wikipedia signed network, it can be observed from Table A3 in Appendix A and Fig. 9 that N144 is ranked as the most negative node by PII measure. It is so because N144 is having all negative connections with 29 nodes including N28, N26, and N80 which are among the well-connected and powerful nodes of the network which makes it fully dependent on other nodes and rendering it powerless. As we trace the curve from left to right in figures, PII is totally deflecting from the Actual curve. From this discussion, it can be analyzed that PII measure is used to compute power of node in political net-

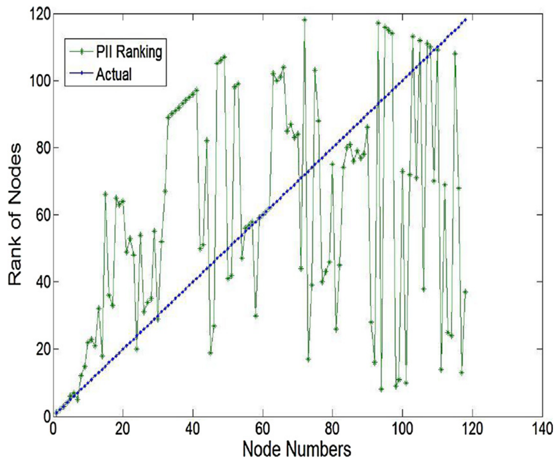


**Fig. 8.** Comparison of PII Ranking versus Actual Ranking in Slashdot signed social network.

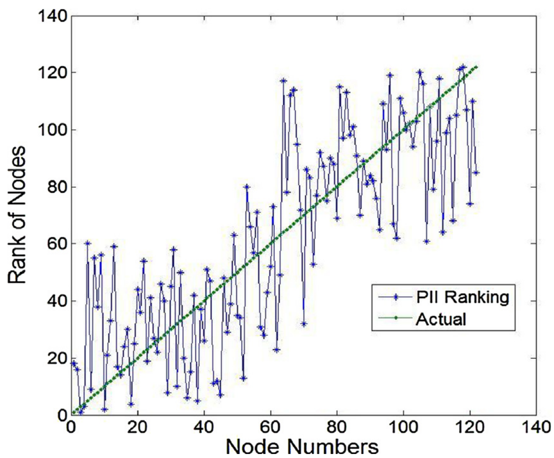


**Fig. 9.** Comparison of PII Ranking versus Actual Ranking in Wikipedia signed social network.

works and its interpretation is completely different from centrality concept.

IV PN Ranking versus Actual Ranking

The interpretation used by this measure to analyze ties is similar to centrality concept of network. In Epinions signed social network, it can be observed from Table A1 in Appendix A that the ranks order given by PN centrality closely matches with the desired ordering. The ranks of first eleven nodes given by PN measure completely overlap with Actual Ranking. As we move down the table, there are very few nodes whose ranks are shuffled. Rest all other ranks are in compliance with the desired order. This can be better analyzed by observing graph of PN versus Actual Ranking in Fig. 10. There are small fluctuations in ranks of PN measure from the Actual curve in the middle of the graph. This happens because here the same value of parameter β is used as it was used in small datasets by [27]. In Slashdot signed social network dataset, the ranking given by PN measure in Slashdot dataset is similar to ranking in Epinions dataset. The first 15 and last 26 ranks are correctly identified by PN centrality, as shown in Table A2 in Appendix A. The middle ranks assigned are very near to the Actual Ranking but does not completely overlap them. It can be confirmed by analyzing Fig. 11. In this dataset, PN measure has identified more number of negative nodes as compared to other measures. In Wikipedia signed social network, it can be analyzed from Fig. 12 that the ranks of nodes are

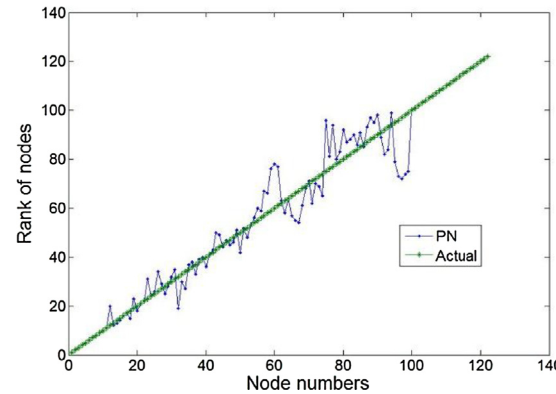


**Fig. 10.** Comparison of PN Ranking versus Actual Ranking in Epinions signed social network dataset.

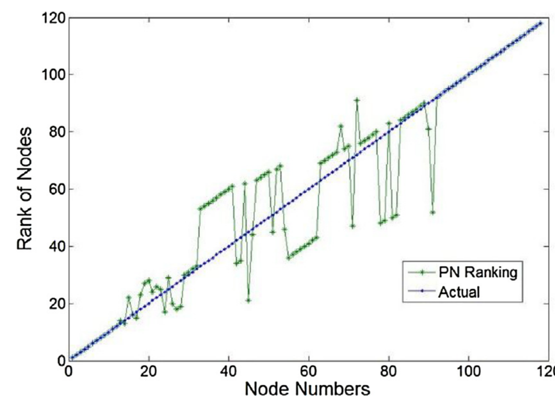


**Fig. 11.** Comparison of PN Ranking versus Actual Ranking in Slashdot signed social network dataset.

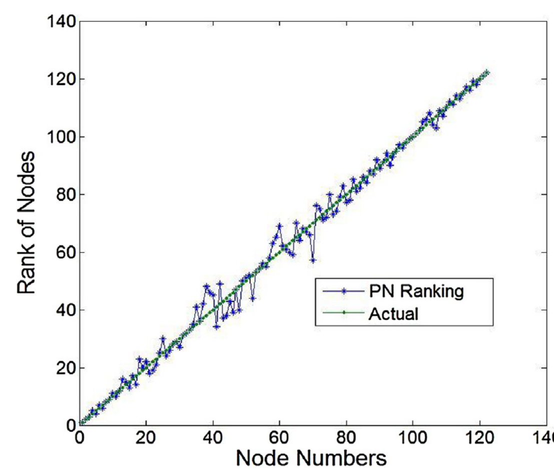


**Fig. 12.** Comparison of PN Ranking versus Actual Ranking in Wikipedia signed social network dataset.

very near to Actual ranking during initial and last few ranks but in the middle, it shows small deviations similar to other two datasets.

From this discussion, it can be deduced that as the network size increases, PN centrality successfully identifies both the most positive and negative nodes of network correctly but cannot identify nodes with intermediary centrality scores. In small networks, PN measure correctly ranks almost all the positive and negative nodes as stated by [27] but in large networks, it shows small variations from Actual Ranking as observed in Figs. 10–12.

V Comparison of all measures using graph

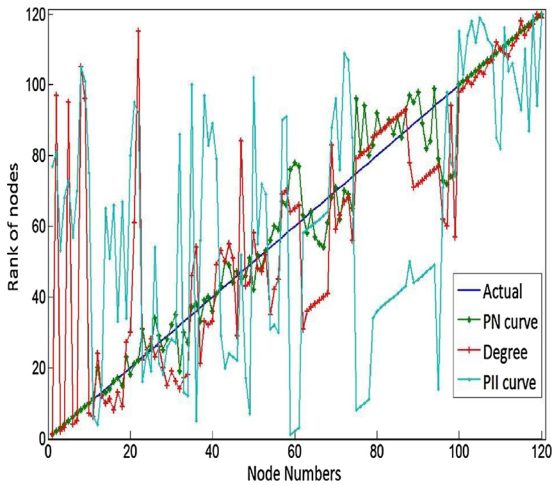


**Fig. 13.** Comparison of all measures versus Actual Ranking in Epinions signed social network dataset.

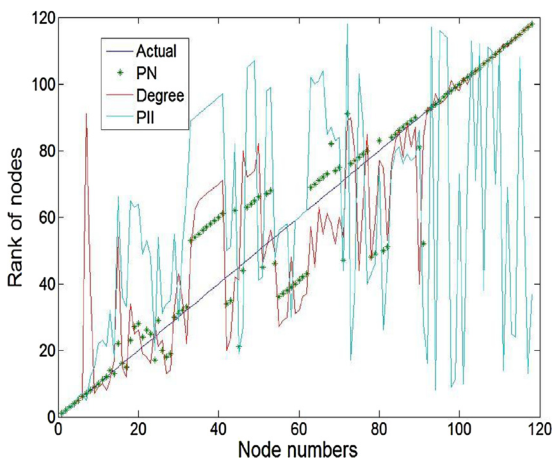


**Fig. 14.** Comparison of all measures versus Actual Ranking in Slashdot signed social network dataset.

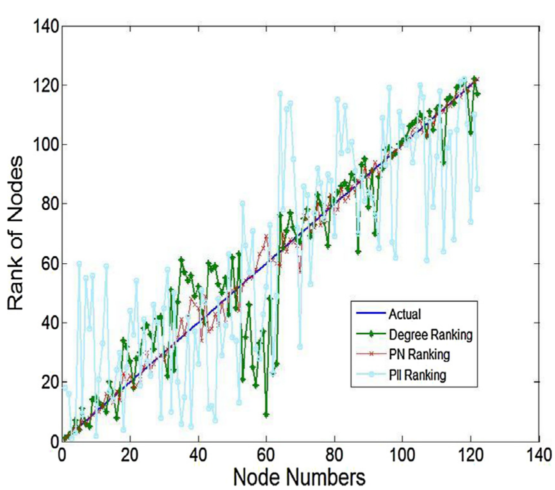


**Fig. 15.** Comparison of all measures versus Actual Ranking in Wikipedia signed social network dataset.

All the three measures are plotted in a single graph along with Actual Ranking to detect the measure which is showing a large amount of deviation from the Actual graph as compared to other measures. In Fig. 13–15, Degree ranking is represented by a red color line, PII by sky blue, PN ranking by Green and Actual Ranking as a straight line in dark blue color. The measures can be compared with each other as well as with Actual Ranking in a better way by analyzing figures.

From Fig. 13–15, following observations are made regarding the performance of mixed data measures in large datasets of online social networks.

i It is observed that all measures performed differently in three different datasets as well as deviates from their behavior that was observed in small datasets.
ii The ranking pattern of PN measure is consistent with regard to Actual Ranking in both the graphs, i.e. it overlaps with an Actual curve at the beginning and at the end of node numbers but shows slight variations in middle node numbers.
iii The ranking pattern obtained from PII measure is different in graphs of two datasets. In Fig. 13, it starts with large deviations from the Actual curve and then it tries to overlap with Actual curve slowly towards the end. In Fig. 14 and 15, it correctly ranks the starting node numbers but then it starts deviating for Actual curve and shows large fluctuations till the end.
iv The ranking pattern of Degree measure is also not consistent in graphs of three datasets. In Epinions dataset, the ranking of degree measure first deviates largely from the Actual curve and then overlaps with it slowly and steadily towards the end of the curve. The same happens in Slashdot and Wikipedia dataset. In starting of the graph, Degree ranked first 7 nodes correctly but then it shows large deviations and while reaching towards the end of the curve it again joins with Actual Ranking.

In nutshell, PN measure being a centrality based index can be used for identifying both the positive and negative nodes in small as well as large networks. There are slight variations from Actual Ranking being observed in the middle of the curve because we have used the older value of parameter β that was used by [27]. In order to get best results from PN measure the value of parameters need to be optimized.

*5.2.5. Optimization*

The results discussed in the previous section proves that in order to get better results from PN measure, the value of parameters used needs to be optimized. Firstly, the reason why the value of parameter β is not able to get good results in larger networks can be found by analyzing its role. The equation of PN centrality can be rewritten as follows:

$$PN = \left[I - \frac{1}{2(n-1)}A\right]^{-1} 1 \qquad (12)$$

The value *1/2(n-1)* used in this equation is parameter $\beta$. This value of $\beta$ is used to normalize positive and negative ties of matrix A, where $n$-1 is the maximum degree of a node in complete network and A = P-2*N. As there are two matrices P and N used in this equation, therefore, $n$-1 is multiplied by 2. The denominator of parameter $\beta$ is the maximum possible degree of a node in a network. This value of $n$-1 in parameter beta is chosen to normalize ties on the basis of fact that maximum boundary condition occurs when the network is having maximum connections, i.e. it is forming a complete graph. When dataset of a large network is carefully observed, it is found that the value of $n$ is very large due to which value of $\beta$ becomes very small and PN centrality scores are as good as degree values of nodes. Also in such a large network, there rarely exists any node which is connected to all other nodes of the network and thus $n$-1 cannot become the maximum possible degree of any node in the network. Therefore, this value of beta does not turn out to be a good normalizing factor.

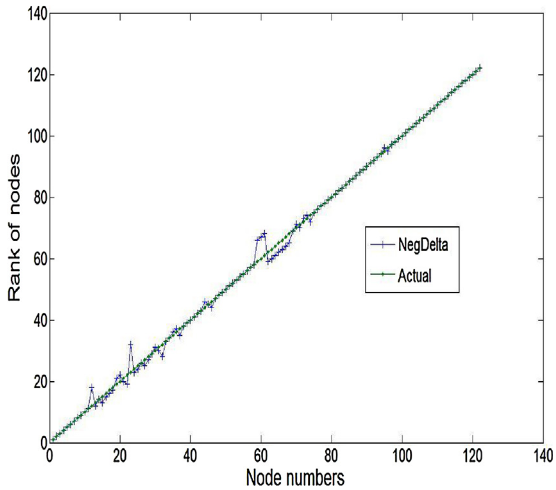


**Fig. 16.** Comparison of $\beta_1$ versus Actual Ranking Epinions signed social network dataset.

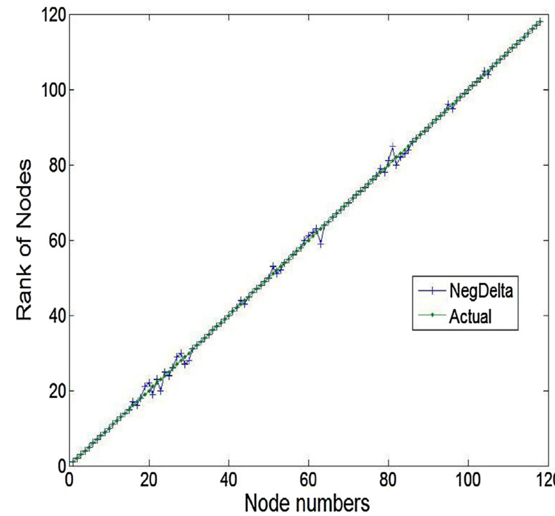


**Fig. 17.** Comparison of $\beta_1$ versus Actual Ranking Slashdot signed social network dataset.

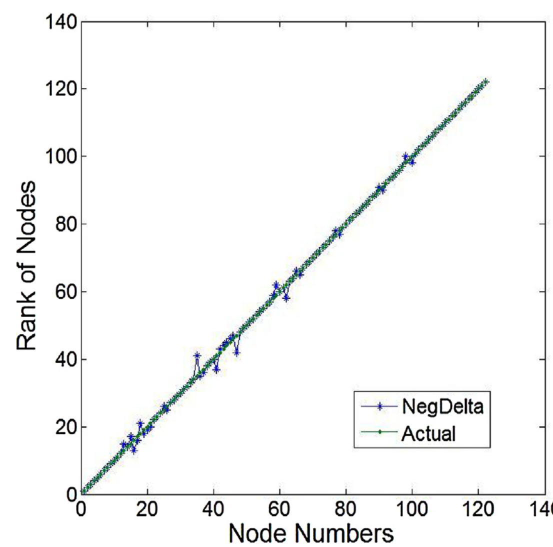


**Fig. 18.** Comparison of $\beta_1$ versus Actual Ranking Wikipedia signed social network dataset.

*5.2.5.1. Projected approach for optimizing results.* For normalizing ties in large networks, some other values of beta need to be chosen that are competent enough to normalize all ties. Many values are suggested by Everett and Borgatti [27] in their paper. One of them is *1/2δ*, where *δ* is the maximum degree in either positive or negative tie matrix. However, for normalization purposes, any value which is substituted for *δ* can better normalize the ties of that particular matrix (P or N) in which *δ* is the maximum possible degree of a node. The other values of beta that are observed from experimental results can be *1/(2*max degree)* and *1/(max degree)*, where (*max degree)* is the maximum possible degree of any node in matrix A which is formed by subtracting 2*N from P. By using these values, the result comes out to be very close to Actual Ranking. The first value of beta i.e. *1/(2*max degree)* performs better than the second one. For further analysis, we have taken different values of beta in three different types of online social networks and visualized their deviations from Actual curve by using graphs. This is discussed as follows:

I Optimizing results in three signed social network datasets

The set of previously taken 122 nodes of Epinions and Wikipedia and 118 nodes of Slashdot is used for analyzing results graphically by optimizing values of parameter β. Three values of β are used for analysis purpose viz. *1/2δ*, where *δ* is the maximum degree in P matrix, *1/2δ* where *δ* is maximum degree in N matrix, *1/(2*max degree)* where (*max degree)* is maximum possible degree in matrix A. The value of β for which curve of PN ranking almost overlaps the Actual curve, i.e. value that identifies maximum number of negative nodes, is regarded as the best value for analyzing ties in larger social network datasets. All the three values of β are analyzed one by one as follows:

a) $\beta_1 = 1/2\delta$ where $\delta$ is maximum degree in 2*N matrix versus Actual Ranking

This value of the beta parameter is used to normalize positive and negative ties of matrix A but it cannot give best results as it can better normalize ties of the negative matrix only. After analyzing Figs. 16–18, it is observed that large fluctuations from Actual curve are mostly depressed and PN ranking almost overlaps the Actual Ranking with very few variations. The result obtained from this value greatly helps to optimize previously obtained results and also identify many negative nodes correctly. But, there are still few deviations present in the middle of this curve as visible in figures.

b) $\beta_2 = 1/2\delta$ where $\delta$ is the maximum degree in P matrix versus Actual Ranking

As *δ* is the maximum degree of a node in positive ties matrix P, so it can better normalize ties of this matrix and overall (i.e. for both P and N) normalization cannot be considered as complete. If we carefully examine the graphs obtained from this value of β in Figs. 19–21, it can be observed that all the deviations become invisible and they completely overlap the Actual curve. Even the variations that are found in Fig. 16–18 are also removed and the smooth straight line is obtained with very fewer points of PN curve away from the Actual curve.

c) $\beta_3$ = *1/(2*maxdegree)* where *maxdegree* is the maximum degree in matrix A versus Actual Ranking

This value of β includes the *maxdegree* variable which is the maximum degree of a node in matrix A. It is to be noted that matrix

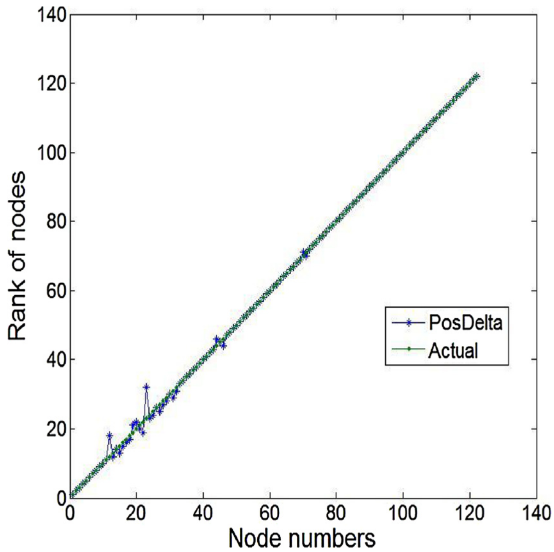


**Fig. 19.** Comparison of $\beta_2$ versus Actual Ranking Epinions signed social network dataset.

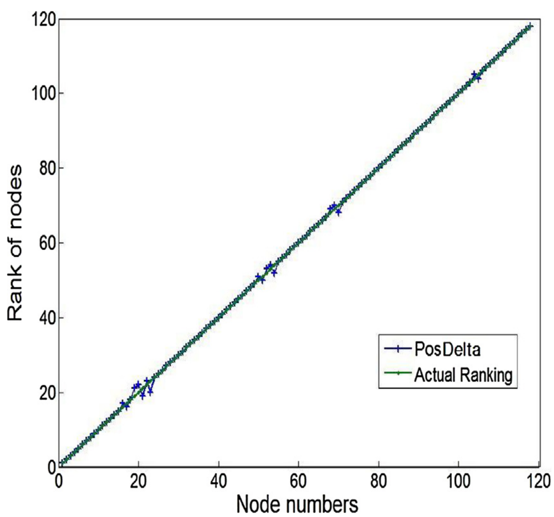


**Fig. 20.** Comparison of $\beta_2$ versus Actual Ranking Slashdot signed social network dataset.

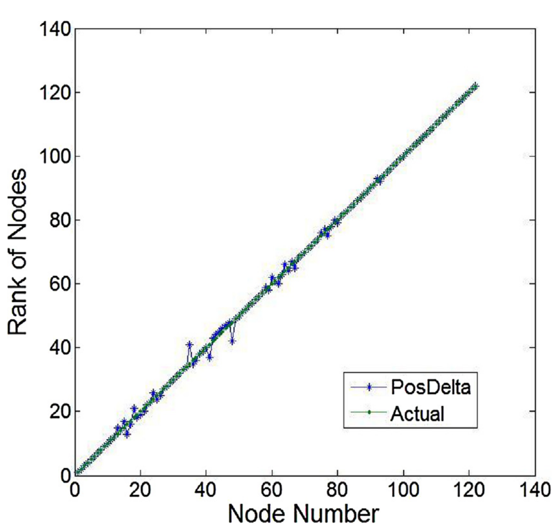


**Fig. 21.** Comparison of $\beta_2$ versus Actual Ranking Wikipedia signed social network dataset.

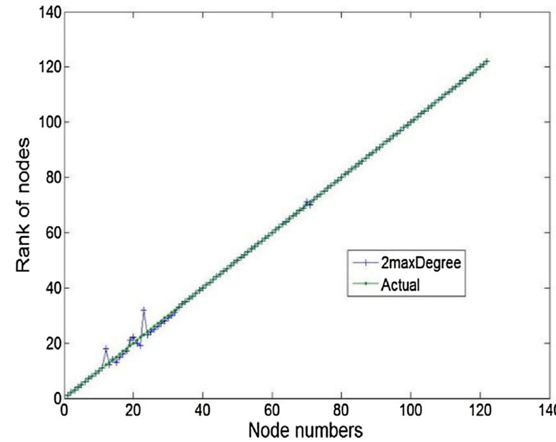


**Fig. 22.** Comparison of $\beta_3$ versus Actual Ranking Epinions signed social network dataset.

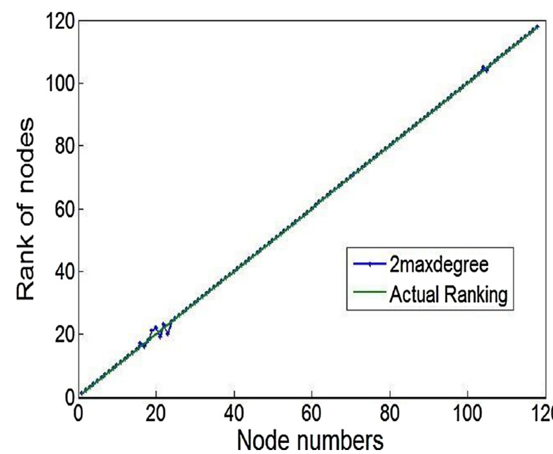


**Fig. 23.** Comparison of $\beta_3$ versus Actual Ranking Slashdot signed social network dataset.

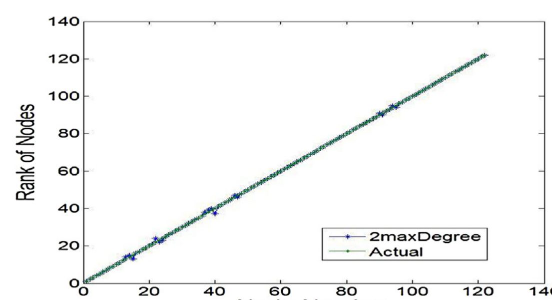


**Fig. 24.** Comparison of $\beta_3$ versus Actual Ranking Wikipedia signed social network dataset.

A is achieved by subtracting twice the negative ties matrix N from positive ties matrix P, i.e. A = P-2*N. It seems conceptually consistent and it can better normalize both the positive and negative ties because *maxdegree* is calculated using both types of ties. The result obtained from this value of beta is presented in Figs. 22–24.

II Comparison of three values of parameter $\beta$ of PN centrality

As it is highlighted in the previous section, there exist three values of $\beta$ for improving the performance of PN centrality measure in large datasets of online social networks. The value of $\beta$, which is able to remove most of the variations in PN ranking curve of Figs. 10–12 of three different datasets, can be identified by analyzing the data of

**Table 1**
Values of parameters in Epinions and Slashdot dataset

| Parameters | Epinions Social Network | Slashdot Social Network | Wikipedia Election Dataset |
|---|---|---|---|
| Maximum degree in P | 970 | 418 | 181 |
| Maximum degree in N matrix & 2*N matrix respectively | 494 988 | 202 404 | 116 232 |
| Maximum degree in A matrix | 962 | 414 | 105 |
| $\beta_1$ value | 0.0005154639 | 0.001196172 | 0.0027624309 |
| $\beta_2$ value | 0.0005060729 | 0.0012376238 | 0.0043103448 |
| $\beta_3$ value | 0.0005197505 | 0.0012077295 | 0.0047619047 |

Table 1. From Figs. 16–24, it can be observed that $\beta_3$ is performing better than other two values of beta in graphs of both the datasets. $\beta_1$ value does not improve the performance of PN measure as compared to other two values. This has happened because it used the maximum degree of N matrix to normalize both the positive and negative ties. As shown in Table 1, the value of *maxdegree* in N is much smaller than other two values in both the datasets.

The $\beta_2$ value also performs well in improving the effectiveness of PN measure but fails to provide more accuracy than $\beta_3$ value. If Figs. 19–24 are carefully analyzed, very little difference between graphs of $\beta_2$ and $\beta_3$ values is observed. Also, the difference between the values of *maxdegree* in P and A matrix is very small. Now, the question arises that which value should be preferred out of these two in order to analyze ties in bigger datasets. Conceptually, the *maxdegree* of $\beta_3$ is able to normalize both the ties equally and *maxdegree* of $\beta_2$ normalizes positive ties better than negative. Coincidentally, the value of a maximum positive degree and maximum total degree is very close to each other in all the datasets that have been taken. Because of this, graphs of both parameters are looking similar. Generally, the overall degree of a node in the network is much smaller than the maximum positive degree of the node. This is because a node can have many negative ties, which, when get subtracted from maximum positive ties results in a decrease of overall degree score. This creates a difference between values of both $\beta_2$ and $\beta_3$ and subsequently, the overall score of each actor also changes.

### 5.2.6. *Calculation of accuracy of measures*

It is clear from above discussion that $\beta = 1/(2^*maxdegree)$ is able to identify most of the negative nodes in large datasets. This result has been ascertained by analyzing graphs of Ranks of nodes vs. Node numbers. However, this result is obtained from the visualization of subsets of nodes in sample datasets of online websites. For detecting the negative nodes in a complete sample of datasets, the graphical techniques become inefficient. Therefore, the statistical methods need to be used for checking the accuracy of ranking given by different measures with respect to Actual Ranking in sample datasets.

*5.2.6.1. Kendall rank correlation coefficient.* This measure is used to compare the ranking given by two variables on the same set of observations. If the ranking given by two variables is highly correlated then the value of correlation is 1 and if it is completely not correlated then the coefficient value is −1. The accuracy of ranking correlation can be verified by following formula:

$$\Gamma = \frac{(\text{no.ofconcordantpairs}) - (\text{no.ofdiscordantpairs})}{n(n-1)/2} \quad (7)$$

To compare the effectiveness of three measures Degree, PN, and PII, the Kendall Tau's coefficient is calculated by above formula. The values of coefficient of each measure are shown in Table 2.

From Table 2, the coefficient of PII measure is lowest as compared to all other measures. It is less than 0.5 in all the three datasets. This happens because of its feature to observe the power of node due to which it cannot correctly identify the centrality of a node in the network. The correlation of Degree measure with Actual Ranking is also very less and hence, cannot be considered as a good measure for identifying ties. From Degree, PII and PN, the correlation of PN measure is quite good but is not acceptable for identifying the most influential as well as the most negative nodes of the network. The Kendall coefficient of different values of $\beta$ in Table 2 shows that $\beta_3$ has outperformed the other values of $\beta$ and attained very high correlation with desired Actual Ranking. For analyzing ties and thus, identifying negative nodes in large datasets of online social networks, a $\beta_3$ parameter of PN measure can be used for obtaining best results.

**Table 2**
Kendall rank correlation coefficient of measures for Epinions, Slashdot, and Wikipedia datasets

| Measure Name | Kendall Rank Correlation Coefficient | | |
|---|---|---|---|
| | Epinions Dataset | Slashdot Dataset | Wikipedia Dataset |
| Degree | 0.6841 | 0.73688 | 0.7613 |
| PII | 0.2817 | 0.2432 | 0.5749 |
| PN | 0.9000 | 0.8572 | 0.8133 |
| $\beta_1$ | 0.9848 | 0.9925 | 0.9911 |
| $\beta_2$ | 0.9927 | 0.9965 | 0.9919 |
| $\beta_3$ | 0.9940 | 0.9980 | 0.9973 |

*5.2.6.2. Chi-square test for goodness-of-fit.* We have also used Chi-Square index for the goodness of fit, to detect how well the ranking given by different measures fits the Actual Ranking. This test is used to assess the association between two measures and identifies the fitness of observed values with the expected model. This statistical test is used to determine the significant level at which the measures are associated with Actual Ranking and is calculated as follows:

$$\chi^2 = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \quad (8)$$

The $\chi^2$ values of different measures are calculated in relation to Actual Ranking at different levels of significance and at given degrees of freedom. The results obtained from Chi-Square test of three different datasets are described in Tables 3–5.

In all the three datasets, it can be observed that there is a large difference between calculated $\chi^2$ values of PII measure with respect to critical values of $\chi^2$ at different levels of significance. Similar results are also obtained for Degree measure. This shows that these measures are inefficient for ranking nodes in large datasets. In case of PN measure, the calculated Chi-Square value is less than critical value at 10% level of significance in Epinions dataset and is greater than all the critical values in Slashdot and Wikipedia datasets. From this observation, it can be inferred that PN measure depicts uncertainty in detecting negative nodes in large datasets at the desired level of significance. Further, three values of $\beta$ proposed to improve the performance of PN exhibits good fit with Actual Ranking at all levels of significance observed in all the datasets. However, the calculated $\chi^2$ value of $\beta_3$ is lowest among other values of parameter $\beta$ in PN measure and hence, produces best results for analyzing negative nodes in online social networks.

*5.2.6.3. F-score test.* We have used another index also named as F-score, to compare the efficiency of different measures in the clas-

**Table 3**
Chi-Square values of different measures in Epinions dataset

| Chi-Square Values | Measures | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| Calculated $\chi^2$ values | 97.3554 | 18218 | 8958.9 | 12.6344 | 8.5760 | 8.26 |
| $\chi^2$ at 1 % significance | 87.7734 | 87.7734 | 87.7734 | 87.7734 | 87.7734 | 87.7734 |
| $\chi^2$ at 2.5% significance | 92.446 | 92.446 | 92.446 | 92.446 | 92.446 | 92.446 |
| $\chi^2$ at 5% significance | 96.5984 | 96.5984 | 96.5984 | 96.5984 | 96.5984 | 96.5984 |
| $\chi^2$ at 10% significance | 101.541 | 101.541 | 101.541 | 101.541 | 101.541 | 101.541 |

**Table 4**
Chi-Square values of different measures in Slashdot dataset

| Chi-Square Values | Measures | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| Calculated $\chi^2$ values | 309.519 | 3221.7 | 1775.9 | 2.6457 | 1.4163 | 1.1782 |
| $\chi^2$ at 1% significance | 84.377 | 84.377 | 84.377 | 84.377 | 84.377 | 84.377 |
| $\chi^2$ at 2% significance | 88.955 | 88.955 | 88.955 | 88.955 | 88.955 | 88.955 |
| $\chi^2$ at 5% significance | 93.0258 | 93.0258 | 93.0258 | 93.0258 | 93.0258 | 93.0258 |
| $\chi^2$ at 10% significance | 97.874 | 97.874 | 97.874 | 97.874 | 97.874 | 97.874 |

**Table 5**
Chi-Square values of different measures in Wikipedia dataset

| Chi-Square Values | Measures | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| Calculated $\chi^2$ values | 394.541 | 2761.3 | 2492.21 | 4.7619 | 4.2371 | 2.0722 |
| $\chi^2$ at 1 % significance | 87.7734 | 87.7734 | 87.7734 | 87.7734 | 87.7734 | 87.7734 |
| $\chi^2$ at 2.5% significance | 92.446 | 92.446 | 92.446 | 92.446 | 92.446 | 92.446 |
| $\chi^2$ at 5% significance | 96.5984 | 96.5984 | 96.5984 | 96.5984 | 96.5984 | 96.5984 |
| $\chi^2$ at 10% significance | 101.541 | 101.541 | 101.541 | 101.541 | 101.541 | 101.541 |

**Table 6**
F-score of different measures in Epinions dataset

| PARAMETERS | MEASURES | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| TRUE POSITIVE | 56/61 | 36/61 | 50/61 | 58/61 | 61/61 | 61/61 |
| FALSE POSITIVE | 5/61 | 25/61 | 11/61 | 3/61 | 0/61 | 0/61 |
| FALSE NEGATIVE | 5/61 | 25/61 | 11/61 | 3/61 | 0/61 | 0/61 |
| F-SCORE | 91.8 | 59.01 | 81.9 | 95.08 | 100 | 100 |

**Table 7**
F-score of different measures in Slashdot dataset

| PARAMETERS | MEASURES | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| TRUE POSITIVE | 36/50 | 27/50 | 35/50 | 48/50 | 49/50 | 50/50 |
| FALSE POSITIVE | 14/68 | 23/68 | 15/68 | 2/68 | 1/68 | 0/68 |
| FALSE NEGATIVE | 14/50 | 23/50 | 15/50 | 2/50 | 1/50 | 0/50 |
| F-SCORE | 76.27 | 61.01 | 74.57 | 96.61 | 98.3 | 100 |

sification of negative nodes of given dataset. The results obtained for F-score of three different datasets are described in Tables 6–8.

The set of 122 nodes in Epinions and Wikipedia datasets is divided into two partitions such that one has only positive nodes and another has only negative nodes. Various parameters such as true positive, false negative and false positive are calculated on the basis of results obtained from different measures. F-score is then calculated using the formula:

$$F_{score} = \frac{2 * \text{TruePositives}}{2 * \text{TruePositives} + \text{FalsePositives} + \text{FalseNegatives}} \tag{11}$$

**Table 8**
F-score of different measures in Wikipedia dataset

| PARAMETERS | MEASURES | | | | | |
|---|---|---|---|---|---|---|
| | PN Measure | PII Measure | DEGREE Measure | $\beta_1$ | $\beta_2$ | $\beta_3$ |
| TRUE POSITIVE | 59/61 | 56/61 | 50/61 | 60/61 | 60/61 | 61/61 |
| FALSE POSITIVE | 5/61 | 5/61 | 11/61 | 1/61 | 1/61 | 0/61 |
| FALSE NEGATIVE | 5/61 | 5/61 | 11/61 | 1/61 | 1/61 | 0/61 |
| F-SCORE | 96.72 | 91.80 | 93.44 | 98.36 | 98.36 | 100 |

The measure with highest F-score is the most efficient that can correctly recognize negative nodes in a given network. From Tables 6 and 8, it can be seen that the most efficient measure is PN because it has a high value of F-score as compared to Degree and PII. Out of different values of parameter β, $\beta_3$ has attained highest F-score and maximum accuracy. Thus, it can be concluded that PN measure with a $\beta_3$ value of parameter produces best results for analyzing negative ties in online social networks.

Similarly, the set of 118 nodes of Slashdot dataset is partitioned into two divisions such that there are 50 negative nodes and 68 positive nodes in the network. The analysis from Table 7 shows that PN measure is better than Degree and PII in identifying negative nodes of the network. The value $\beta_3$ has outperformed the other two values of i.e. $\beta_1$ and $\beta_2$ and thus, $\beta_3$ can be used for identifying negative nodes in large datasets of online social networks.

The results obtained from all the three statistical techniques validates that PN performed better than Degree and PII measures but its accuracy of finding ranks in bigger networks is low and not acceptable. The comparison of the performance of three proposed values of parameter β by rank correlation, chi-square statistic, and F-score shows that the ranking given by $\beta_3$ is more correlated with Actual Ranking than the other two values. Thus, it can be concluded that PN measure with $\beta_3$ value is the most efficient measure of ranking nodes and thereby, identify negative nodes more accurately.

## 6. Summary and conclusion

In this paper, the problem of analyzing negative nodes in online social networks has been addressed. In today's era, where social networks are online and contain millions of nodes, the analysis of fraudulent activities becomes the need of the hour. Various network analysis measures such as Degree, Status, PII, and PN have been proposed by researchers to analyze negative nodes on small datasets of social networks. However, no attempts were made to recognize the negative nodes in large online social networks. In this paper, an approach is proposed to rank each node in the network on the basis of centrality possessed by it in order to identify negative nodes of the network. Evaluation of performance of each measure in identifying strangers (i.e. negatively behaving nodes) in the online networks has been carried out by using the statistical and graphical approaches.

After analyzing the obtained results, it is concluded that PN centrality is able to identify more number of negative nodes as compared to any other measure but still shows some variations from Actual Ranking. So, for attaining more accurate results in case of large networks, three different values of parameter β (i.e. normalization factor) have also been evaluated and it has been observed that $\beta_3 = 1/(2^*\text{max degree in A matrix})$ produces the best results. The research in the current area can be extended by using the concept of negative cliques (i.e. smallest group in which everyone outside the group dislikes someone within the group) to identify negative nodes. Apart from this, the application of PN centrality measure in different online social networks can further be improved by exploring more values of β. Moreover, curve fitting approaches can be used for identifying outsiders of the network by taking PN centrality as a graph metric.

## Appendix A. The name and rank of the specific set of nodes given by Degree, PII and PN measures in Epinions, Slashdot and Wikipedia signed social network datasets in Tables A1–A3 respectively

**Table A1**
Results of Epinions signed social network dataset.

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N79 | 1 | 1 | 1 | 77 | 1 |
| N5 | 2 | 2 | 97 | 81 | 2 |
| N23 | 3 | 3 | 2 | 53 | 3 |
| N25 | 4 | 4 | 3 | 68 | 4 |
| N52 | 5 | 5 | 95 | 73 | 5 |
| N106 | 6 | 6 | 4 | 57 | 6 |
| N31 | 7 | 7 | 5 | 70 | 7 |
| N39 | 8 | 8 | 105 | 105 | 8 |
| N41 | 9 | 9 | 96 | 101 | 9 |
| N113 | 10 | 10 | 7 | 75 | 10 |
| N410 | 11 | 11 | 6 | 6 | 11 |
| N37772 | 12 | 12 | 24 | 4 | 20 |
| N15505 | 13 | 13 | 12 | 15 | 12 |
| N9310 | 14 | 14 | 10 | 65 | 13 |
| N14124 | 15 | 15 | 11 | 51 | 14 |
| N865 | 16 | 16 | 8 | 66 | 16 |
| N112779 | 17 | 17 | 13 | 33 | 17 |
| N2475 | 18 | 18 | 9 | 67 | 15 |
| N24238 | 19 | 19 | 27 | 34 | 23 |
| N1652 | 20 | 20 | 30 | 80 | 18 |
| N79 | 21 | 21 | 61 | 95 | 21 |
| N57 | 22 | 22 | 115 | 92 | 22 |
| N19064 | 23 | 23 | 22 | 16 | 31 |
| N44421 | 24 | 24 | 25 | 25 | 24 |
| N2395 | 25 | 25 | 28 | 19 | 26 |
| N12047 | 26 | 26 | 23 | 54 | 34 |
| N20491 | 27 | 27 | 26 | 21 | 29 |
| N116770 | 28 | 28 | 20 | 18 | 25 |
| N1853 | 29 | 29 | 15 | 26 | 28 |
| N92799 | 30 | 30 | 19 | 28 | 32 |
| N17391 | 31 | 31 | 16 | 27 | 35 |
| N264 | 32 | 32 | 14 | 86 | 19 |
| N57053 | 33 | 33 | 17 | 13 | 30 |
| N58764 | 34 | 34 | 18 | 12 | 27 |
| N666 | 35 | 35 | 46 | 100 | 37 |
| N66 | 36 | 36 | 54 | 5 | 38 |
| N1532 | 37 | 37 | 21 | 56 | 33 |
| N8684 | 38 | 38 | 33 | 97 | 39 |
| N6176 | 39 | 39 | 32 | 83 | 40 |
| N6296 | 40 | 40 | 33 | 89 | 36 |
| N11349 | 41 | 41 | 49 | 79 | 41 |
| N55 | 42 | 42 | 53 | 29 | 43 |
| N7576 | 43 | 43 | 50 | 20 | 50 |
| N81 | 44 | 44 | 55 | 24 | 49 |
| N42 | 45 | 45 | 51 | 23 | 44 |
| N1 | 46 | 46 | 29 | 22 | 47 |
| N108 | 47 | 47 | 84 | 52 | 45 |
| N6 | 48 | 48 | 43 | 17 | 46 |
| N37720 | 49 | 49 | 44 | 7 | 51 |
| N1118 | 50 | 50 | 58 | 102 | 42 |
| N26240 | 51 | 51 | 48 | 55 | 52 |
| N6271 | 52 | 52 | 47 | 72 | 48 |
| N21940 | 53 | 53 | 52 | 69 | 53 |
| N14416 | 54 | 54 | 35 | 31 | 56 |
| N100958 | 55 | 55 | 42 | 32 | 60 |
| N1127 | 56 | 56 | 45 | 30 | 59 |
| N943 | 57 | 57 | 69 | 90 | 67 |
| N1120 | 58 | 58 | 70 | 91 | 66 |
| N42175 | 59 | 59 | 64 | 1 | 76 |
| N47153 | 60 | 60 | 65 | 2 | 78 |
| N51867 | 61 | 61 | 66 | 3 | 77 |
| N4590 | 62 | 62 | 31 | 58 | 63 |
| N16560 | 63 | 63 | 36 | 59 | 58 |
| N20383 | 64 | 64 | 37 | 60 | 64 |
| N22466 | 65 | 65 | 38 | 61 | 57 |
| N25564 | 66 | 66 | 39 | 62 | 55 |
| N59755 | 67 | 67 | 40 | 63 | 54 |
| N66466 | 68 | 68 | 41 | 64 | 61 |
| N5591 | 69 | 69 | 83 | 88 | 68 |
| N1119 | 70 | 70 | 59 | 96 | 71 |
| N3253 | 71 | 71 | 63 | 76 | 62 |
| N270 | 72 | 72 | 67 | 109 | 70 |
| N7047 | 73 | 73 | 68 | 107 | 69 |
| N558 | 74 | 74 | 56 | 85 | 65 |
| N49 | 75 | 75 | 79 | 8 | 96 |

Table A1 (*Continued*)

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N14136 | 76 | 76 | 80 | 9 | 81 |
| N37436 | 77 | 77 | 81 | 10 | 94 |
| N60411 | 78 | 78 | 82 | 11 | 80 |
| N115 | 79 | 79 | 85 | 34 | 83 |
| N910 | 80 | 80 | 86 | 36 | 92 |
| N15774 | 81 | 81 | 87 | 37 | 87 |
| N34858 | 82 | 82 | 88 | 38 | 88 |
| N48372 | 83 | 83 | 89 | 39 | 90 |
| N54858 | 84 | 84 | 90 | 40 | 86 |
| N67647 | 85 | 85 | 91 | 41 | 91 |
| N112298 | 86 | 86 | 92 | 42 | 85 |
| N113248 | 87 | 87 | 93 | 43 | 93 |
| N47 | 88 | 88 | 78 | 50 | 97 |
| N12193 | 89 | 89 | 71 | 44 | 95 |
| N27905 | 90 | 90 | 72 | 45 | 98 |
| N35807 | 91 | 91 | 73 | 46 | 89 |
| N38967 | 92 | 92 | 74 | 47 | 82 |
| N52491 | 93 | 93 | 75 | 48 | 84 |
| N59959 | 94 | 94 | 76 | 49 | 99 |
| N21739 | 95 | 95 | 77 | 14 | 79 |
| N515 | 96 | 96 | 62 | 71 | 73 |
| N3144 | 97 | 97 | 60 | 98 | 72 |
| N8778 | 98 | 98 | 94 | 78 | 74 |
| N1003 | 99 | 99 | 57 | 74 | 75 |
| N2293 | 100 | 100 | 98 | 115 | 100 |
| N363 | 101 | 101 | 99 | 103 | 101 |
| N86 | 102 | 102 | 101 | 114 | 102 |
| N433 | 103 | 103 | 100 | 118 | 103 |
| N75 | 104 | 104 | 102 | 112 | 104 |
| N83 | 105 | 105 | 104 | 119 | 105 |
| N63 | 106 | 106 | 103 | 117 | 106 |
| N122 | 107 | 107 | 106 | 113 | 107 |
| N93 | 108 | 108 | 107 | 111 | 108 |
| N102 | 109 | 109 | 112 | 85 | 109 |
| N87 | 110 | 110 | 110 | 82 | 110 |
| N59 | 111 | 111 | 109 | 116 | 111 |
| N14 | 112 | 112 | 108 | 104 | 112 |
| N21 | 113 | 113 | 111 | 106 | 113 |
| N88 | 114 | 114 | 113 | 99 | 114 |
| N35 | 115 | 115 | 118 | 93 | 115 |
| N56 | 116 | 116 | 114 | 110 | 116 |
| N48 | 117 | 117 | 116 | 87 | 117 |
| N97 | 118 | 118 | 117 | 121 | 118 |
| N116 | 119 | 119 | 120 | 94 | 119 |
| N104 | 120 | 120 | 119 | 120 | 120 |
| N50 | 121 | 121 | 121 | 108 | 121 |
| N20 | 122 | 122 | 122 | 122 | 122 |

**A**: Names of Nodes, **B**: Node numbers, **C**: Rank of nodes, **D**: Degree ranking, **E**: PII ranking, **F**: PN ranking

**Table A2**
Results of Slashdot signed social network dataset

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N46 | 1 | 1 | 1 | 1 | 1 |
| N48 | 2 | 2 | 2 | 2 | 2 |
| N137 | 3 | 3 | 3 | 3 | 3 |
| N103 | 4 | 4 | 4 | 4 | 4 |
| N22 | 5 | 5 | 5 | 6 | 5 |
| N19 | 6 | 6 | 6 | 7 | 6 |
| N34 | 7 | 7 | 91 | 5 | 7 |
| N107 | 8 | 8 | 39 | 12 | 8 |
| N1810 | 9 | 9 | 7 | 15 | 9 |
| N1808 | 10 | 10 | 9 | 22 | 10 |
| N5022 | 11 | 11 | 10 | 23 | 11 |
| N676 | 12 | 12 | 8 | 21 | 12 |
| N23 | 13 | 13 | 11 | 32 | 14 |
| N118 | 14 | 14 | 17 | 18 | 13 |
| N11052 | 15 | 15 | 54 | 66 | 22 |
| N6035 | 16 | 16 | 15 | 36 | 16 |
| N1824 | 17 | 17 | 12 | 33 | 15 |
| N7570 | 18 | 18 | 34 | 65 | 23 |
| N2477 | 19 | 19 | 25 | 63 | 27 |
| N2478 | 20 | 20 | 26 | 64 | 28 |
| N750 | 21 | 21 | 19 | 49 | 24 |
| N172 | 22 | 22 | 18 | 53 | 26 |

Table A2 (*Continued*)

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N25 | 23 | 23 | 16 | 48 | 25 |
| N2523 | 24 | 24 | 28 | 20 | 17 |
| N1785 | 25 | 25 | 21 | 54 | 29 |
| N1825 | 26 | 26 | 23 | 31 | 20 |
| N2502 | 27 | 27 | 13 | 34 | 18 |
| N2553 | 28 | 28 | 14 | 35 | 19 |
| N6517 | 29 | 29 | 33 | 55 | 30 |
| N1240 | 30 | 30 | 43 | 29 | 31 |
| N8302 | 31 | 31 | 35 | 52 | 32 |
| N1803 | 32 | 32 | 22 | 67 | 33 |
| N3094 | 33 | 33 | 50 | 89 | 53 |
| N10204 | 34 | 34 | 62 | 90 | 54 |
| N10938 | 35 | 35 | 65 | 91 | 55 |
| N11047 | 36 | 36 | 66 | 92 | 56 |
| N11048 | 37 | 37 | 67 | 93 | 57 |
| N11049 | 38 | 38 | 68 | 94 | 58 |
| N11050 | 39 | 39 | 69 | 95 | 59 |
| N11051 | 40 | 40 | 70 | 96 | 60 |
| N11053 | 41 | 41 | 71 | 97 | 61 |
| N1338 | 42 | 42 | 20 | 50 | 34 |
| N1999 | 43 | 43 | 24 | 51 | 35 |
| N1075 | 44 | 44 | 42 | 82 | 62 |
| N632 | 45 | 45 | 41 | 19 | 21 |
| N1287 | 46 | 46 | 80 | 27 | 44 |
| N12189 | 47 | 47 | 72 | 105 | 63 |
| N12968 | 48 | 48 | 73 | 106 | 64 |
| N17849 | 49 | 49 | 74 | 107 | 65 |
| N1981 | 50 | 50 | 82 | 41 | 66 |
| N1442 | 51 | 51 | 46 | 42 | 45 |
| N3296 | 52 | 52 | 51 | 98 | 67 |
| N6631 | 53 | 53 | 56 | 99 | 68 |
| N2569 | 54 | 54 | 49 | 47 | 46 |
| N2520 | 55 | 55 | 27 | 56 | 36 |
| N2538 | 56 | 56 | 29 | 57 | 37 |
| N2542 | 57 | 57 | 30 | 58 | 38 |
| N1801 | 58 | 58 | 48 | 30 | 39 |
| N4546 | 59 | 59 | 31 | 59 | 40 |
| N5391 | 60 | 60 | 32 | 60 | 41 |
| N10760 | 61 | 61 | 36 | 61 | 42 |
| N10765 | 62 | 62 | 37 | 62 | 43 |
| N6767 | 63 | 63 | 57 | 102 | 69 |
| N1428 | 64 | 64 | 45 | 100 | 70 |
| N10584 | 65 | 65 | 63 | 101 | 71 |
| N6587 | 66 | 66 | 55 | 104 | 72 |
| N8347 | 67 | 67 | 61 | 85 | 73 |
| N7605 | 68 | 68 | 58 | 87 | 82 |
| N3972 | 69 | 69 | 52 | 83 | 74 |
| N8337 | 70 | 70 | 60 | 84 | 75 |
| N6586 | 71 | 71 | 54 | 44 | 47 |
| N8311 | 72 | 72 | 89 | 118 | 91 |
| N59 | 73 | 73 | 90 | 17 | 76 |
| N1265 | 74 | 74 | 79 | 39 | 77 |
| N1364 | 75 | 75 | 44 | 103 | 78 |
| N10864 | 76 | 76 | 64 | 88 | 79 |
| N3974 | 77 | 77 | 85 | 40 | 80 |
| N1650 | 78 | 78 | 47 | 43 | 48 |
| N7614 | 79 | 79 | 59 | 46 | 49 |
| N1084 | 80 | 80 | 77 | 75 | 83 |
| N670 | 81 | 81 | 75 | 26 | 50 |
| N4791 | 82 | 82 | 53 | 45 | 51 |
| N748 | 83 | 83 | 76 | 74 | 84 |
| N2113 | 84 | 84 | 83 | 80 | 85 |
| N4344 | 85 | 85 | 86 | 81 | 86 |
| N1248 | 86 | 86 | 78 | 76 | 87 |
| N10907 | 87 | 87 | 88 | 79 | 88 |
| N1520 | 88 | 88 | 81 | 77 | 89 |
| N7401 | 89 | 89 | 87 | 78 | 90 |
| N209 | 90 | 90 | 40 | 86 | 81 |
| N2514 | 91 | 91 | 84 | 28 | 52 |
| N4 | 92 | 92 | 93 | 16 | 92 |
| N71 | 93 | 93 | 92 | 117 | 93 |
| N63 | 94 | 94 | 97 | 8 | 94 |
| N55 | 95 | 95 | 94 | 116 | 95 |
| N92 | 96 | 96 | 95 | 115 | 96 |
| N61 | 97 | 97 | 96 | 114 | 97 |
| N57 | 98 | 98 | 101 | 9 | 98 |
| N14 | 99 | 99 | 99 | 11 | 99 |

Table A2 (*Continued*)

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N95 | 100 | 100 | 98 | 73 | 100 |
| N78 | 101 | 101 | 102 | 10 | 101 |
| N65 | 102 | 102 | 100 | 72 | 102 |
| N91 | 103 | 103 | 103 | 113 | 103 |
| N66 | 104 | 104 | 104 | 71 | 104 |
| N106 | 105 | 105 | 105 | 112 | 105 |
| N97 | 106 | 106 | 106 | 38 | 106 |
| N67 | 107 | 107 | 107 | 111 | 107 |
| N70 | 108 | 108 | 108 | 110 | 108 |
| N80 | 109 | 109 | 109 | 70 | 109 |
| N94 | 110 | 110 | 110 | 109 | 110 |
| N64 | 111 | 111 | 112 | 14 | 111 |
| N83 | 112 | 112 | 111 | 69 | 112 |
| N142 | 113 | 113 | 113 | 25 | 113 |
| N74 | 114 | 114 | 114 | 24 | 114 |
| N87 | 115 | 115 | 115 | 108 | 115 |
| N8 | 116 | 116 | 116 | 68 | 116 |
| N62 | 117 | 117 | 117 | 13 | 117 |
| N82 | 118 | 118 | 118 | 37 | 118 |

**A**: Names of Nodes, **B**: Node numbers, **C**: Rank of nodes, **D**: Degree ranking, **E**: PII ranking, **F**: PN ranking

**Table A3**
Results of Wikipedia signed social network dataset

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N2354 | 1 | 1 | 1 | 77 | 1 |
| N94 | 2 | 2 | 2 | 81 | 2 |
| N144 | 3 | 3 | 3 | 53 | 3 |
| N906 | 4 | 4 | 5 | 68 | 7 |
| N192 | 5 | 5 | 4 | 73 | 4 |
| N875 | 6 | 6 | 7 | 57 | 11 |
| N477 | 7 | 7 | 6 | 70 | 6 |
| N3243 | 8 | 8 | 8 | 105 | 5 |
| N81 | 9 | 9 | 9 | 101 | 14 |
| N5467 | 10 | 10 | 11 | 75 | 15 |
| N5415 | 11 | 11 | 10 | 6 | 13 |
| N682 | 12 | 12 | 12 | 4 | 12 |
| N302 | 13 | 13 | 16 | 15 | 10 |
| N741 | 14 | 14 | 15 | 65 | 20 |
| N5387 | 15 | 15 | 13 | 51 | 16 |
| N829 | 16 | 16 | 17 | 66 | 8 |
| N368 | 17 | 17 | 14 | 33 | 17 |
| N5465 | 18 | 18 | 23 | 67 | 34 |
| N938 | 19 | 19 | 20 | 34 | 32 |
| N5326 | 20 | 20 | 22 | 80 | 27 |
| N3258 | 21 | 21 | 18 | 95 | 18 |
| N32 | 22 | 22 | 19 | 92 | 28 |
| N5255 | 23 | 23 | 21 | 16 | 30 |
| N167 | 24 | 24 | 25 | 25 | 38 |
| N934 | 25 | 25 | 30 | 19 | 39 |
| N3808 | 26 | 26 | 24 | 54 | 36 |
| N435 | 27 | 27 | 26 | 21 | 31 |
| N515 | 28 | 28 | 28 | 18 | 41 |
| N930 | 29 | 29 | 29 | 26 | 42 |
| N514 | 30 | 30 | 27 | 28 | 29 |
| N2785 | 31 | 31 | 31 | 27 | 22 |
| N525 | 32 | 32 | 32 | 86 | 51 |
| N5123 | 33 | 33 | 33 | 13 | 24 |
| N5457 | 34 | 34 | 35 | 12 | 47 |
| N2535 | 35 | 35 | 41 | 100 | 61 |
| N463 | 36 | 36 | 42 | 5 | 57 |
| N4338 | 37 | 37 | 48 | 56 | 54 |
| N5470 | 38 | 38 | 46 | 97 | 56 |
| N5417 | 39 | 39 | 45 | 83 | 49 |
| N5393 | 40 | 40 | 34 | 89 | 52 |
| N3976 | 41 | 41 | 49 | 79 | 44 |
| N4 | 42 | 42 | 37 | 29 | 40 |
| N904 | 43 | 43 | 38 | 11 | 60 |
| N903 | 44 | 44 | 43 | 12 | 58 |
| N5351 | 45 | 45 | 39 | 7 | 59 |
| N3255 | 46 | 46 | 47 | 48 | 53 |
| N5426 | 47 | 47 | 40 | 29 | 50 |
| N4212 | 48 | 48 | 50 | 39 | 55 |
| N93 | 49 | 49 | 51 | 63 | 43 |

Table A3 (*Continued*)

| A | B | C | D | E | F |
|---|---|---|---|---|---|
| N121 | 50 | 50 | 52 | 35 | 62 |
| N4468 | 51 | 51 | 44 | 34 | 45 |
| N5354 | 52 | 52 | 53 | 13 | 63 |
| N95 | 53 | 53 | 54 | 80 | 21 |
| N243 | 54 | 54 | 56 | 66 | 35 |
| N5375 | 55 | 55 | 55 | 57 | 46 |
| N5323 | 56 | 56 | 58 | 71 | 25 |
| N5341 | 57 | 57 | 63 | 31 | 19 |
| N110 | 58 | 58 | 65 | 28 | 33 |
| N79 | 59 | 59 | 69 | 43 | 37 |
| N214 | 60 | 60 | 62 | 52 | 9 |
| N5020 | 61 | 61 | 61 | 73 | 48 |
| N25 | 62 | 62 | 60 | 23 | 23 |
| N299 | 63 | 63 | 59 | 49 | 26 |
| N2116 | 64 | 64 | 70 | 117 | 76 |
| N608 | 65 | 65 | 64 | 78 | 65 |
| N348 | 66 | 66 | 68 | 112 | 71 |
| N560 | 67 | 67 | 67 | 114 | 77 |
| N370 | 68 | 68 | 66 | 95 | 72 |
| N50 | 69 | 69 | 57 | 72 | 68 |
| N20 | 70 | 70 | 76 | 32 | 67 |
| N1247 | 71 | 71 | 75 | 86 | 75 |
| N402 | 72 | 72 | 71 | 83 | 78 |
| N9 | 73 | 73 | 72 | 53 | 69 |
| N33 | 74 | 74 | 80 | 77 | 73 |
| N1393 | 75 | 75 | 73 | 92 | 83 |
| N19 | 76 | 76 | 74 | 87 | 80 |
| N178 | 77 | 77 | 79 | 75 | 74 |
| N856 | 78 | 78 | 83 | 90 | 66 |
| N5463 | 79 | 79 | 77 | 88 | 82 |
| N339 | 80 | 80 | 78 | 69 | 81 |
| N54 | 81 | 81 | 85 | 115 | 84 |
| N3460 | 82 | 82 | 81 | 97 | 86 |
| N228 | 83 | 83 | 82 | 113 | 87 |
| N895 | 84 | 84 | 82 | 98 | 85 |
| N2713 | 85 | 85 | 86 | 101 | 90 |
| N733 | 86 | 86 | 84 | 91 | 88 |
| N4712 | 87 | 87 | 88 | 70 | 64 |
| N308 | 88 | 88 | 87 | 89 | 93 |
| N4929 | 89 | 89 | 92 | 81 | 95 |
| N75 | 90 | 90 | 89 | 84 | 79 |
| N5155 | 91 | 91 | 91 | 82 | 91 |
| N5449 | 92 | 92 | 94 | 76 | 70 |
| N80 | 93 | 93 | 90 | 65 | 89 |
| N2909 | 94 | 94 | 93 | 109 | 92 |
| N5430 | 95 | 95 | 95 | 93 | 98 |
| N3251 | 96 | 96 | 97 | 119 | 99 |
| N26 | 97 | 97 | 96 | 67 | 96 |
| N29 | 98 | 98 | 98 | 62 | 97 |
| N817 | 99 | 99 | 99 | 111 | 100 |
| N3103 | 100 | 100 | 100 | 106 | 101 |
| N633 | 101 | 101 | 101 | 100 | 102 |
| N439 | 102 | 102 | 102 | 102 | 106 |
| N1814 | 103 | 103 | 105 | 94 | 107 |
| N2822 | 104 | 104 | 106 | 103 | 108 |
| N1799 | 105 | 105 | 108 | 120 | 110 |
| N36 | 106 | 106 | 104 | 116 | 109 |
| N11 | 107 | 107 | 103 | 61 | 103 |
| N3910 | 108 | 108 | 109 | 108 | 111 |
| N55 | 109 | 109 | 107 | 79 | 105 |
| N3027 | 110 | 110 | 110 | 96 | 112 |
| N2135 | 111 | 111 | 112 | 118 | 113 |
| N8 | 112 | 112 | 111 | 64 | 94 |
| N3253 | 113 | 113 | 114 | 99 | 115 |
| N4448 | 114 | 114 | 113 | 104 | 116 |
| N6 | 115 | 115 | 115 | 68 | 114 |
| N4600 | 116 | 116 | 117 | 105 | 119 |
| N2225 | 117 | 117 | 116 | 121 | 120 |
| N4400 | 118 | 118 | 119 | 122 | 121 |
| N86 | 119 | 119 | 118 | 107 | 118 |
| N28 | 120 | 120 | 120 | 74 | 104 |
| N271 | 121 | 121 | 121 | 110 | 122 |
| N5412 | 122 | 122 | 122 | 85 | 122 |

**A**: Names of Nodes, **B**: Node numbers, **C**: Rank of nodes, **D**: Degree ranking, **E**: PII ranking, **F**: PN ranking

## References


[1] F. Heider, Attitudes and cognitive organization, J. Psychol. 21 (1) (1946) 107–112 (Taylor & Francis Group).
[2] D. Cartwright, F. Harary, Structural balance: a generalization of Heider's theory, Psychol. Rev. 63 (no. 5) (1956).
[3] J.A. Davis, Clustering and Structural Balance in Graphs, Human Relations, Sage Publications, 1967.
[4] R.L. Breiger, S.A. Boorman, P. Arabie, An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling, J. Math. Psychol. 12 (1975) 328–383.
[5] J.P. Boyd, Social Semi-groups: A Unified Theory of Scaling and Blockmodelingas Applied to Social Networks, George Mason University Press, Fairfax,VA, 1990.
[6] F.J. Roethlisberger, W.J. Dickson, Management and the Worker, Cambridge University Press, Cambridge, 2003.
[7] K. Read, Cultures of the Central Highlands, New Guinea, Southwestern Journal of Anthropology, University of New Mexico and the Laboratory of Anthropology, 1954, pp. 1–43.
[8] S. Sampson, Crisis in a Cloister, CornellUniversity, 1967 (Unpublished Ph.D. Thesis).
[9] R.V. Guha, R. Kumar, P. Raghavan, A. Tomkins, Propagation of trust and distrust, Proc. 13th WWW (2004).
[10] P. Massa, P. Avesani, Controversial Users Demand Local Trust Metrics: An Experimental Study on epinions.com Community, AAAI '05, AAAI Press, 2005, pp. 121–126.
[11] M.J. Brzozowski, T. Hogg, G. Szabó, Friends and foes: ideological social networking, Proc. 26th CHI (2008).
[12] J. Kunegis, A. Lommatzsch, C. Bauckhage, The Slashdot Zoo: mining a social network with negative edges, Proc. 18th WWW (2009) 741–750.
[13] C. Lampe, E. Johnston, P. Resnick, Follow the reader: filtering comments on slashdot, Proc. 25th CHI (2007).
[14] B. Viswanath, M.A. Bashir, M. Crovella, S. Guha, K.P. Gummadi, B. Krishnamurthy, A. Mislove, Towards detecting anomalous user behavior in online social networks, Proceedings of the 23rd USENIX Security Symposium (USENIX Security) (2014) (August).
[15] L. Akoglu, H. Tong, D. Koutra, Graph based anomaly detection and description: a survey, Data Mining Knowl. Discov. 29 (3) (2014) 626–688.
[16] Y. Liu, S. Chawla, Social media anomaly detection: challenges and solutions, Proceedings of the 21 st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2015) 2317–2318 (August).
[17] R. Yu, X. He, Y. Liu, GLAD: group anomaly detection in social media analysis, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2014) 372–381 (August).
[18] D. Savage, X. Zhang, X. Yu, P. Chou, Q. Wang, Anomaly detection in online social networks, Soc. Netw. 39 (2014) 62–70.
[19] Q. Cheng, X. Lu, Z. Liu, J. Huang, Mining research trends with anomaly detection models: the case of social computing research, Scientometrics 103 (2) (2015) 453–469.
[20] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, N.F. Samatova, Anomaly detection in dynamic networks: a survey, Wiley Interdiscip. Rev.: Comput. Stat. 7 (3) (2015) 223–247.
[21] L.C. Freeman, Centrality in social networks conceptual clarification, Social Netw. 1 (3) (1979) 215–239.
[22] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43, Springer-Verlag.
[23] C.H. Hubbell, An input-output approach to clique identification, Sociometry (1965) 377–399, American Sociological Association.
[24] P. Bonacich, Power and centrality: a family of measures, in: American Journal of Sociology, University of Chicago Press, 1987, pp. 1170–1182.
[25] P. Bonacich, P. Lloyd, Calculating status with negative relations, Soc. Netw. 26 (4) (2004) 331–338.
[26] J.M. Smith, D.S. Halgin, V. Kidwell-Lopez, G. Labianca, D.J. Brass, S.P. Borgatti, Power in politically charged networks, Soc. Netw. 36 (2014) 162–176.
[27] M.G. Everett, S.P. Borgatti, Networks containing negative ties Social Networks, vol. 38, North-Holland, 2014, pp. 111–120.
[28] F. Lorrain, H.C. White, Structural equivalence of individuals in social networks, J. Math. Sociol. 1 (1) (1971) 49–80.
[29] D.R. White, K.P. Reitz, Graph and semigroup homomorphisms on networks of relations, Soc. Netw. vol. 5 (no. 2) (1983) 193–234 (North-Holland).
[30] M.G. Everett, S.P. Borgatti, Regular equivalence: general theory, J. Math. Sociol. 19 (1) (1994) 29–52.
[31] L. Hubert, J. Schultz, Quadratic assignment as a general data analysis strategy, Br. J. Math. Stat. Psychol. 29 (2) (1976) 190–241 (Blackwell Publishing Ltd.).
[32] D. Dekker, D. Krackhardt, T.A.B. Snijders, Sensitivity of MRQAP tests to collinearity and autocorrelation conditions, Psychometrika 72 (4) (2007) 563–581.
[33] S.P. Borgatti, Centrality and AIDS, Connections 18 (1) (1995) 112–114.
[34] S.P. Borgatti, Centrality and network flow, Soc. Netw. 27 (1) (2005) 55–71.
[35] G. Barbian, Trust centrality in online social networks, in: Proc. of the European Intelligence and Security Informatics Conference, EISIC 2011, 2017, pp. 372–377 (Athens).
[36] P.D. Meo, K. Musial-Gabrys, D. Rosaci, G.M. Sarnè, L. Aroyo, Using centrality measures to predict helpfulness-based reputation in trust networks, CM Trans. Internet Technol. (TOIT) 17 (1) (2017).
[37] I. Varlamis, M. Eirinaki, M. Louta, A study on social network metrics and their application in trust networks, in: Proc. of the International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2010, 2010, pp. 168–175 (2010, August).
[38] J. Leskovec, D. Huttenlocher, J. Kleinberg, Signed networks in social media, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2010) 1361–1370.
[39] J. Leskovec, D. Huttenlocher, J. Kleinberg, Predicting positive and negative links in online social networks, Proceedings of the 19th International Conference on World Wide Web (2010) 641–650.
[40] J. Leskovec, R. Sosič, Snap: stanford network analysis platform, ACM Trans. Intell. Syst. Technol.(TIST) 8 (1) (2016).
[41] J. Leskovec, C. Faloutsos, "Sampling from large graphs.", Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2006) 631–636, August.


**Mankirat Kaur** received M.E. degree in Computer Science and Engineering from University Institute of Engineering and Technology, Panjab University, Chandigarh, India. She received her B.Tech degree in Computer Science and Engineering from Punjabi University, Patiala, Punjab, Indiain 2013. Her research interests include Social network analysis and classification of social data.

**Sarbjeet Singh** received B. hisTech degree in Computer Science and Engineering from Punjab Technical University, Jallandhar, Punjab, Indiain 2001 and M.E. degree is Software Engineering from Thapar University, Patiala, India in 2003. He also received Ph.D degree in Computer Science and Engineering from Thapar University, Patiala, India in 2009, working on grid security systems architecture. Currently he is Associate Professor in Computer Science and Engineering at UIET, Panjab University, Chandigarh, India. He has more than 30 research publications in international journals and conferences to his credit. His research interests include parallel and distributed systems, distributed security architectures, distributed services like grid and web services, privacy and trust related issues in distributed environments Dr. Singh is a life member of Computer Society of India and Indian Society for Technical Education.