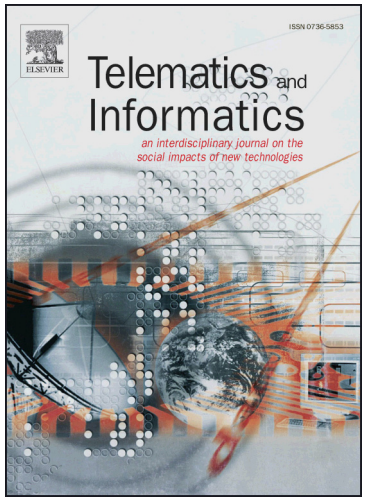


A Trust Model for Analysis of Trust, Influence and their Relationship in Social Network Communities

Yousra Asim, Ahmad Kamran Malik, Basit Raza, Ahmad Raza Shahid

PII: S0736-5853(18)30895-5
DOI: https://doi.org/10.1016/j.tele.2018.11.008
Reference: TELE 1192

To appear in: *Telematics and Informatics*

Received Date: 9 August 2018
Revised Date: 19 October 2018
Accepted Date: 20 November 2018

Please cite this article as: Asim, Y., Malik, A.K., Raza, B., Shahid, A.R., A Trust Model for Analysis of Trust, Influence and their Relationship in Social Network Communities, *Telematics and Informatics* (2018), doi: https://doi.org/10.1016/j.tele.2018.11.008

A Trust Model for Analysis of Trust, Influence and their Relationship in Social Network Communities

Yousra Asim, Ahmad Kamran Malik*, Basit Raza, Ahmad Raza Shahid

Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan.

engryousraasim@gmail.com, ahmad.kamran@comsats.edu.pk, basit.raza@comsats.edu.pk, ahmadrshahid@comsats.edu.pk

*Corresponding author: Ahmad Kamran Malik

'Declarations of interest: none'

Acknowledgments

This work has been funded and supported by COMSATS University Islamabad (CUI), Islamabad, Pakistan under research and development, CUI/ORIC-PD/18.

Abstract

Influential nodes are capable of influencing the Social Network (SN) structure and functions. Influential nodes are not pre-defined and need to be identified to better understand and control the network. Trust of the nodes can be an important indicator of their influence and trusted nodes can also affect other nodes. However, analysis of trust and influence relationship in social network communities needs investigation. This study proposes a *SNTrust* model to find the trust of nodes in a network using a local and global trust and also investigates trust, influence and their relationship in SN communities. Different SN-based influence evaluation approaches named K-core, closeness centrality, eigenvector centrality, and page rank is used to find influential nodes. We have explored the trust and influence of nodes in their communities as well as in the network. Different standard datasets such as Consulting Company dataset, Freeman EIES dataset, Blogcatalogue dataset, and Facebook groups dataset are used for experimentation and Pearson's correlation and level of significance (p-value) are used for results evaluation. We found a positive linear correlation between the local trust of a node and its influence. It is found that the nodes which are trusted in the network are highly trusted in their communities. There is a strong linear relationship between the influence of a node in the network and community. Furthermore, it is also observed that nodes that are close to each other in a community have high trust among them. The results show that the proposed *SNTrust* model performed better in finding trust, influence and their correlation.

Keywords: *Trust: Influence: Social network: Community*

1. Introduction

Social Network (SN) is a collection of different types of social actors, their relationships and interactions. All users in a SN do not behave alike and have differences in their cultural, demographic, educational and professional characteristics. Also, they have different relationships of varying strengths with different people. Likewise, not all the users can have the equal capacity to influence others. To put it another way, the degree of importance of each node can vary in the network (Yang and Xie, 2016). Some users are more influential and have a greater impact on other users. Influential nodes are rare but effective to influence others in a SN (Zhao et al., 2017). Their opinions and actions can motivate people to follow them and to adapt their preferences. Such nodes are needed to control the SN (Yang and Xie, 2016), to spread influence to a larger audience (Zhao et al., 2016) and to build relationships (Zhang et al., 2010), for viral marketing, political campaigning, and brand advertisements (Araujo et al., 2017).

Identification of influential nodes among all nodes in the SN is an interesting and intensively studied domain (Yang and Xie, 2016). It has been found that finding the most influential users for influence maximization is an NP-hard problem (Kempe et al., 2003; Wang et al., 2010). However, several techniques have been presented for identification of influential nodes as mentioned in section 3. In SN, users build relationships with each other. A user's opinions can influence others in the decision-making process so influence relationships are significant to highlight important nodes. For example, an opinion of the "best friend" is considered more valuable to a person as compared to a "friend" relationship. The frequency of communication is also usually high in close friends other than casual relationships. Such SN data can be mined to analyze the trust of a person on another.

Trust has different interpretations, and different representations which are widely used in different fields, such as Philosophy, Economics, Psychology, International Relations, Sociology, and Computer science (Cho et al., 2015). It is a multidimensional metric which represents consent of a trustor to have confidence in a trustee for his positive behavior. Trustor and trustee are two major entities of trust where a trustor is a person who evaluates how much he/she credits a trustee whereas a person who is being evaluated by trustor is termed as trustee (Baek and Kim, 2014). In SN, a number of nodes are part of the network, but not all nodes are equally significant or reliable. Their degree of importance is not predefined and it cannot be apparently estimated whether they are trusted or not. To avail the benefits of influential nodes, their credibility is also crucial. Most of the time, SN users come up with a situation where they have to accept friend request from an anonymous person or they have to follow an unknown person without any prior experience and knowledge about him/her. In the case of online shoppings, customers also have to think about the trustworthiness of the seller (Hamdi et al., 2016). Such situations enhance the need of a trust computation method which may help SN users to decide whether to proceed with another party or not.

Thus, detection of important nodes and assuming that they can positively influence others is not sufficient. It is also essential to find whether influential nodes in the network are also trustworthy (can positively influence others or not). Finding such nodes is essential to effectively influence other nodes (Xu et al., 2012). Many trust related studies have highlighted that trust can play a vital role in the decision-making process (Caverlee et al., 2010; Hargittai et al., 2010; Resnick and Zeckhauser, 2002). Social trust is presented as a variable of influence. If trusted nodes are used to spread influence in the network then the number of influenced nodes in the network can be maximized (Yap and Lim, 2017). Different methods have been proposed to find the trust of nodes in SNs (Sherchan et al., 2013). Existing approaches for influential node identification use structural properties of a node (as shown in Fig. 1). Also, some studies that calculate trust of influential nodes in product review domain are available (Liu et al., 2015; Xu et al., 2012).

Though, it has been observed (Xu et al., 2014; Yap and Lim, 2017) that trust of a node positively contributes towards its influential capabilities, however, the question is how they are related to each other? In social networks, trust of nodes may be an important indicator of influence, but it also gives rise to many questions addressing the influence power of these trusted nodes in both the network and in their communities. To the best of our knowledge, rarely some research is performed by measuring, analyzing, and comparing the significance of a node using both trust and influence. Similarly, the statistical association between both of aforesaid metrics has not been previously investigated on the network as well as community level. With this in mind, this research has mainly four objectives. The first objective is to find the trust of nodes and their influence in the SN. The second objective concentrates on finding the relationship between trust and influence of a node. The third objective is to investigate whether highly trusted nodes of the network are also trusted in their own communities. The last objective is to examine whether influential nodes of the network also influential in their own communities.

In the real world, trust and influence both are interdependent. Investigating this relationship with respect to a social network has high significance and relevancy. For instance, if any, new Cosmetics Company wants to advertise its products effectively, it may opt for an online campaign. One possible way out in this context is to select

any well-known blogger as an ambassador and initiator for online advertisement of company products who has more direct connections (number of followers). Another possibility is that not only to choose any well-known blogger in terms of his number of followers, but also to pick out a fashion blogger or beauty blogger who is more relevant in the context of the campaign. The third possibility is that company may give preference to a blogger who is also famous in the communities/groups; he has joined online including the prior options. As the people of similar interests and backgrounds usually become a part of an online community. If a blogger is famous in his community, then his concern is more likely to be rapidly welcomed by community members. The persons related to the fashion industry, showbiz industry, and beauty conscious females might be the part of that community which can make online campaign more successful in a while. It indicates that, not only the influence of an individual is important on the network level for inspiring others, but also his influence in communities does matter. Besides, if the selected influential blogger is also reliable at the same time and people trust him more, then he can be more useful in this online campaign for company advertisement. Because it is natural, we are more likely to be influenced by the person we trust more. Trust in a person can provide grounds for his successful relationships with others. This way, he can effectively and positively contribute to the decision making process of consumers in the society and ultimately can provide significant help to the concerned company. In such scenarios, it seems essential to investigate the relationship between trust and influence on global network and community level.

This research theoretically contributes to the existing pool of knowledge by exploring the statistical connection between trust and influence. The obtained results (as discussed in section 5) point out that trusted nodes can be influential due to the positive linear relationship between both variables. Furthermore, it is found that a strong positive association exists between trusts and influence on the full network level and community level. The relationships under investigation are found statistically significant in all cases. Trust is found high among globally prominent nodes (nodes having high eigenvector centralities) of the communities and the communities having high trust among their members are found denser and more transitive. Moreover, this work empirically offers a trust computation method to determine the trusted and non-trusted nodes in different social network communities and in the full network.

SNTrust incorporates direct trust and the indirect trust of nodes. Different SN data characteristics such as user attributes, degree prestige, the frequency of communication, and relationship types are used for direct trust calculations. The indirect calculation of trust is performed by focusing on different user activities and responses to those activities. Moreover, different well-known SN-based influence evaluating approaches are used to identify influential nodes in the network instead of using one specific method. Subsequently, the comparative analysis of a node's trust and influence is performed at community and network level by carrying out extensive experiments on standard datasets to achieve research objectives. Pearson's correlation and level of significance are used to analyze the results for determining the statistical relationship between trust and influence. Pearson's method not only indicates the presence or absence of correlation between any two variables, but also, determines the exact extent, or degree to which they are correlated, also if the correlation is positive or negative. Experimental results align with the assumptions made while addressing each research objective.

The methodological contribution of this research is that it incorporates the trust of a user in communities instead of only focusing the user's direct connections which makes SNTrust more practical in the real-world scenarios. Users tend to join different online communities along with making online friends while using SN platforms. This work is a medley of various network characteristics in terms of direct trust calculations and indirect trust calculations. The methodology opted for finding the answers of research questions being studied is novel in itself, e.g. community detection to find the trusted and influential nodes on community-level. The proposed SNTrust model is practically useful for Facebook which is a well-known online social network platform. On Facebook, a user can have many types of relationships such as friends, and close friends. The attributes of a user on his Facebook profile such as education, workplace, and location, etc. can be used to find his/her attribute trust with his friends. A user can be a member of many Facebook groups which enables him/her to find the indirect trust of a user. A user can post in these groups against which other group members give their responses such as shares, comments, likes, haha, waooo, and love, etc. that can help to find the response trust of a user. Number of posts of a user in his joined groups can be used to compute his participation trust.

In this paper, Section 2 represents a theoretical background of trust. Section 3 consists of related work. Section 4 presents *SNTrust* model. Section 5 describes datasets, results and their analysis. Section 6 concludes this paper.

2. Theoretical Background of Trust

Firstly, to compute trust, the understanding of the fundamental idea of trust, its types and its properties is unavoidable. Keeping this in mind, the definition of trust, trust types and some used trust properties in SNTrust model are discussed.

2.1 Trust definition

Trust measure indicates the presence of an association among trustor and trustee. As the notion of trust has been diversely used in literature, but in this research work, the definition of trust in the field of computer science as presented in (Jøsang et al., 2007) is considered. It elaborates trust as a subjective measure in which a given person A can have his own opinion about another person B to consider the latter trusted or not.

2.2 Types of Trust

In the context of trust computation, we have considered direct and indirect trust of a user. *Direct trust* illustrates the direct exposure of a trustor in terms of direct interactions with a trustee (Hamdi et al., 2012). If both trustor and trustee have interest similarity, direct interactions, and relationships, then it can contribute to build up the trust relationship between them. Trust has many types such as Calculative, Relational, Emotional, Cognitive, Institutional, and Dispositional (Sherchan et al., 2013), however, in this research work, the following types of trust are focused:

2.2.1 Relational Trust

The trust evolves between the trustor and trustee, if they collaborate and communicate with each other time and again. If they have more positive mutual exchanges, then trust relationships strengthens (Resnick et al., 2000), otherwise diminishes, which shows that the trust is relational.

2.2.2 Emotional Trust

The trust is considered as an emotional measure which makes trustor comfortable to have confidence in trustee (Kuan and Bock, 2005). This emotional aspect of trust forces the trustor to continue the relationship with trustee with positive attitude. Besides, (Taylor, 2000) states that if trustor has former direct experiences with trustee then it can contribute towards his emotional trust on the trustee. The proposed SNTrust model fits in the aforesaid theoretical concepts of trust types. In proposed model, conversational trust (which is a component of direct trust) is based on the *relational* aspect of trust which depicts the frequency of communication between two users. Besides, participation trust (which is a component of indirect trust) is also based on relational aspect of trust which is basically the communication of a person on a group level. Higher the number of posts by user in a group contributes towards his trust relationship with his group members. In the same way, SNTrust incorporates relationship trust and prestige trust which are based on *emotional* perspective of trust. Relationship trust is the true representative of the fact that if a user A has natural relationship e.g. brother of another user B then A can rely on B with full comfort which denotes his emotional trust by being his sibling. Likewise, if a user B has direct positive interactions with another user A then it will add up to user A's prestige trust with respect to user B

2.3 Trust properties

Trust can have different properties trust such as Context-specific, Dynamic, Propagative, Aggregative, Non-Transitive, Subjective, Composable, Asymmetric, Self-Reinforcing, Transitive, Event-based (Sherchan et al., 2013). In SNTrust model, the following trust properties are focused for trust computation of a user:

2.3.1 Subjective

Due to the subjective nature of trust, John can think differently about the Prank's review/comments about a book. Such viewpoints can directly affect the trust value computation.

2.3.2 Asymmetric

Trust has asymmetric nature in the sense that if user A trusts user B then it is not compulsion that vice versa would be true. It occurs due to different beliefs and perceptions of different users.

2.3.3 Non-Transitive

Trust is non-transitive in nature, which indicates that if a user A trusts user B, and user B trust user C, then user A trusts user C is not true.

2.3.4 Context-specific

This property of trust says that a trustor can trust on trustee in one context and may not be in the other context.

2.3.5 Composable

Trust has composable property. It specifies that a user can form some trust on another user even if he is not directly connect to him due to propagation of trust along social chains.

The suggested SNTrust model conforms to the preceding trust properties. In the perspective of *subjective* property of trust, the average attribute trust value of a user A is calculated by taking into account all attribute trust values given by his neighbors. Because each neighbor of user A can have different attributes with respect to the former which leads to varying attribute trust values. Likewise, in the calculation of relationship trust of user A, each of his neighbors can have varying perceptions for different relationships. For example, Alice can be the close friend of Bob, but can be only a friend of Prank at the same time. What's more, Alice's post can get shared by Bob (for being closer to Bob) but it is likely that it gets only liked by Prank, which makes the nature of response trust as subjective. The computation of trust in this work is *disproportionate*, because trust value between two users depends upon many network properties e.g. attributes, relationship etc. which may vary from user to user. It results different values of trust between two users with respect to each other. The SNTrust model is *non-transitive* in nature which shows that it has no property like if Bob trusts Naina and Naina trusts Prank, then Bob trusts Prank.

In the proposed model, attribute trust calculation follows *context-specific* property of the trust. For example, Alice trusts Bob as an engineer, but not as a good cook. Bob is trustworthy in the context of being an engineer with respect to Alice, but he is not trustworthy in cooking good food. The response trust calculation in the suggested model is based on the comments, likes, shares received on a post by a user to calculate his trust from the perspective of other group members. It is likely to follow *composable* trust property. Moreover, in case of a number of social chains recommending trust for a user, trust information from all chains is composed for trust computation of that user. In the context of attribute trust calculation in the proposed model, the local network (alters) of a user (ego) is considered only and then trust information is composed to obtain the average trust value of that user.

3. Related work

Detection of influential nodes is challenging because there is no clear definition of being influential (Sheikhahmadi et al., 2017). To highlight right influential nodes is difficult (Probst et al., 2013) due to a large number of users in the network (Sheikhahmadi et al., 2017) and the activities and behavior of users also vary (Khadangi and Bagheri, 2017). The influence of a node in the network can be determined by using different methods proposed in the literature. These methods use global and/or local structure information to evaluate the influence of a node. For example, closeness centrality of a node is measured with respect to the whole network (i.e. average length of distance of a node from all other nodes in the network) (Kiss and Bichler, 2008; Sabidussi, 1966). Other examples of influential nodes detection methods based on global measures are: Betweenness (Kiss and Bichler, 2008; Kitsak et al., 2010), K-core/k-shell (Kitsak et al., 2010), Eigenvector centrality (Bonacich and Lloyd, 2001; Kiss and Bichler, 2008), PageRank (Heidemann et al., 2010), and Leader-Rank (Lü et al., 2011). Degree centrality of a node is measured by using its local structure information (number of adjacent nodes) (Kiss and Bichler, 2008; Probst et al., 2013).

Some of the influential node detection methods based on local measures include Local Structural Centrality Measure (Gao et al., 2014), H-index (Lü et al., 2016), Degree Distance (Sheikhahmadi et al., 2015), Edge weighted degree centrality (Kiss and Bichler, 2008), Structural Hole (Burt, 1993), and ClusterRank (Chen et al., 2013). Additionally, some hybrid methods which use both local and global information of a node to calculate its influence include Coefficient of Local Centrality (Zhao et al., 2017), Weighted Leader Rank (Li et al., 2014), INN (Sheikhahmadi et al., 2017), and Coreness Centrality (Bae and Kim, 2014). Methods using global information for influence identification have high computational costs because of a large number of nodes in the network whereas methods using local information of a node are less effective in the sense that they only consider the neighborhood of a node. Positions of nodes and combination of multiple methods are also used for ranking nodes in the network (Wang et al., 2017). The Fig. 1 categorizes the available influential node identification methods into global, local, and hybrid methods.

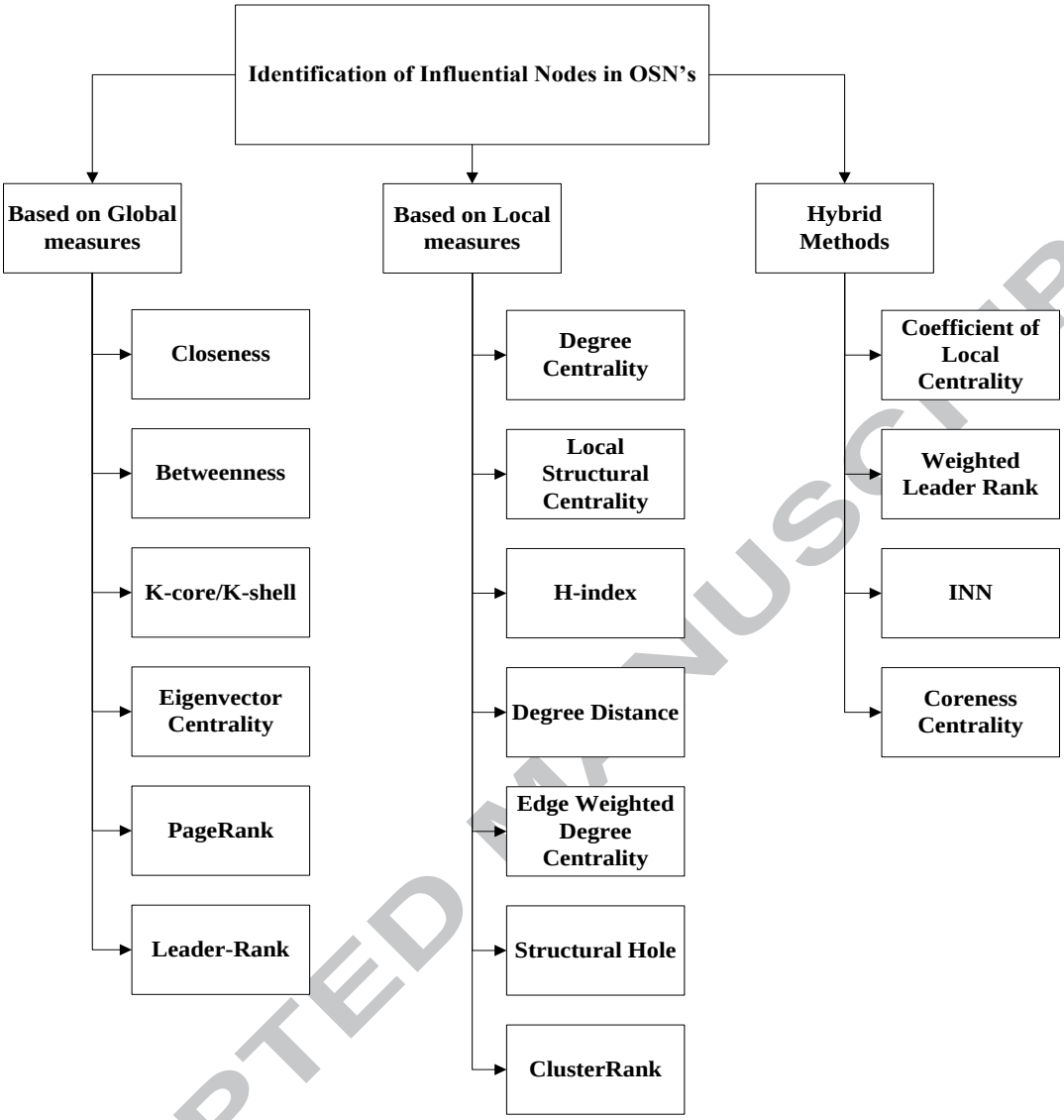


Fig. 1: Taxonomy of methods proposed for Influential Nodes Identification

Trustworthiness of nodes is related to influence (Xu et al., 2012). Trust can play an important role in a user's decision-making (Caverlee et al., 2010; Hargittai et al., 2010; Resnick and Zeckhauser, 2002). A number of trust calculation methods exist for SNs (Sherchan et al., 2013). These methods can be classified into three categories: *network-structure based methods*, *user-interactions based methods*, and *hybrid methods*. Moreover, trust and influence are jointly explored in a few studies as well, for example, an approach is presented to identify the valuable customers in the network for profit maximization (Xu et al., 2012). In this study, the authors used positive and negative rating interactions and trust lists of users to propose semi-definite programming optimization technique targeting enterprises. This technique outperforms other benchmark methods. Product review domain-aware approach to identify effective influential nodes using trust, domain and time are presented in (Liu et al., 2015). Results show that proposed approach is more accurate as compared to both degree centrality and popular author algorithm. IRIS is a method that can calculate direct trust value between users (Hamdi et al., 2012). It considers positive and negative interactions between users, interest similarity and user relationships for trust computation. It categories social network users in three groups: malicious, benevolent and controversial users. Indirect trust is not focused in this approach. Although many studies are available to find the importance of a node by calculating its trust and influence in SNs, however, the relationship between both metrics (trust and influence) on the network and community level is under-explored.

4. SNTrust model

This section describes the details of *SNTrust* model. It includes calculation of trust using many trust components at different levels. *SNTrust* includes trust at macro and micro level. *SNTrust* is able to provide trust at three levels including user, community, and network level. To achieve the research objectives of this study, first of all, we need algorithms to find trusted nodes in the network. This paper presents a novel *SNTrust* model to find the trust of nodes in the network using *Direct Trust* and *Indirect Trust*. Further, in this research, we have used well-known influential node identification approaches named *K-core, Closeness, Page Rank,* and *Eigenvector* to find the importance/influence of a node. For this purpose, we have opt for global measures instead of local measures because former measures show good performance in node ranking by focusing a node's position in the whole network (Wang et al., 2017). Alternately, local measures are less effective in the sense that they only consider the local neighborhood (local contacts) based on a node's location by ignoring global structural information of the network. Beside, hybrid measures are not used in this work due to their complex mechanisms to find influence of a node. Moreover, it is previously mentioned that K-core is capable of achieving exceptional performance with O(n) complexity which makes it feasible to be applied on larger networks (Wang et al., 2017). Closeness centrality can identify the most influential nodes with a minimum number of intercessors (Sheikhahmadi et al., 2015). PageRank calculates a node's influence by considering the direction and associated weights of its links with other powerful nodes of the network (Asim et al., 2018). It is a variant of eigenvector centrality which ignores the direction of links in calculating influence of a node by considering the influence of other connected most influential nodes with it.

Four standard datasets are used for evaluation. The standard datasets of a Consulting company, *Freeman EIES dataset, and Blogcatalog dataset* are used to find the node's importance and comparative analysis of both approaches. A real-world dataset of three Facebook groups is selected to find trusted nodes using *Indirect Trust*. *Pearson's correlation* and *level of significance (p-value)* are used for results evaluation. The Fig. 2 shows the steps followed in this paper to achieve our research goals.

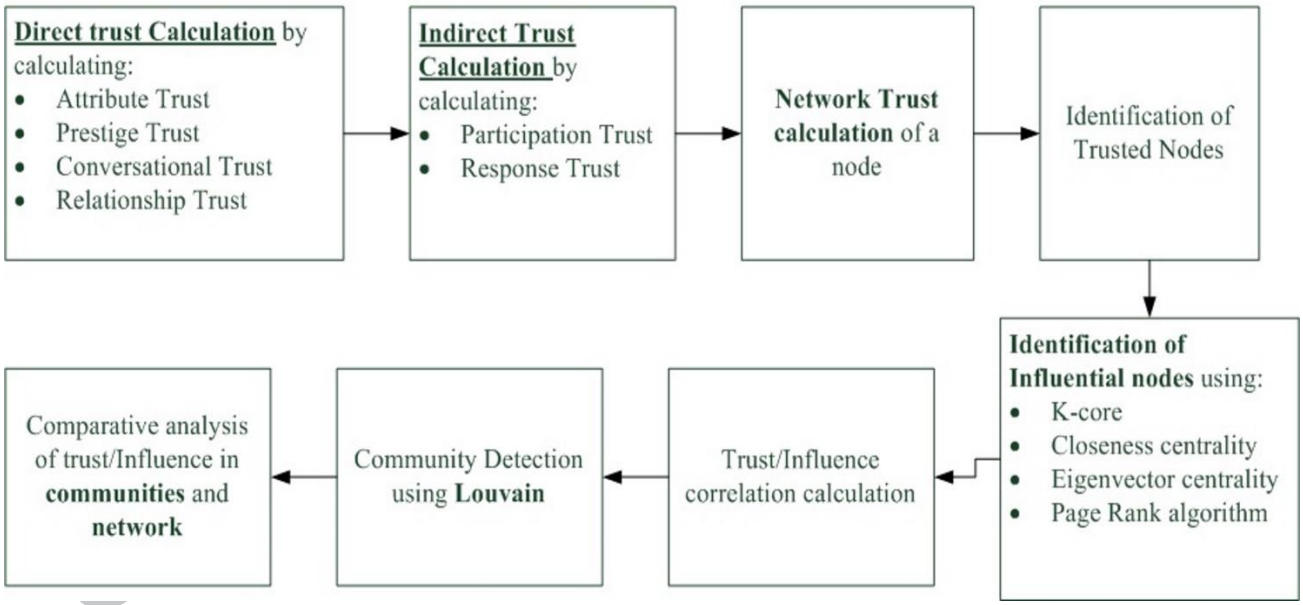


Fig. 2: A proposed methodology for trust and influence in social network communities

To achieve our first research goal, we propose a method to calculate trust of a node. SN data of a node can be used in determining its importance, instead of just exploring typical structural measures of node importance. SN characteristics such that node attributes, relationship types, relationship strengths, node degree as prestige, and node activities are considered for determining trusted nodes in the network. The Fig. 3 given below shows the bottom up approach followed to find the trust of a user.

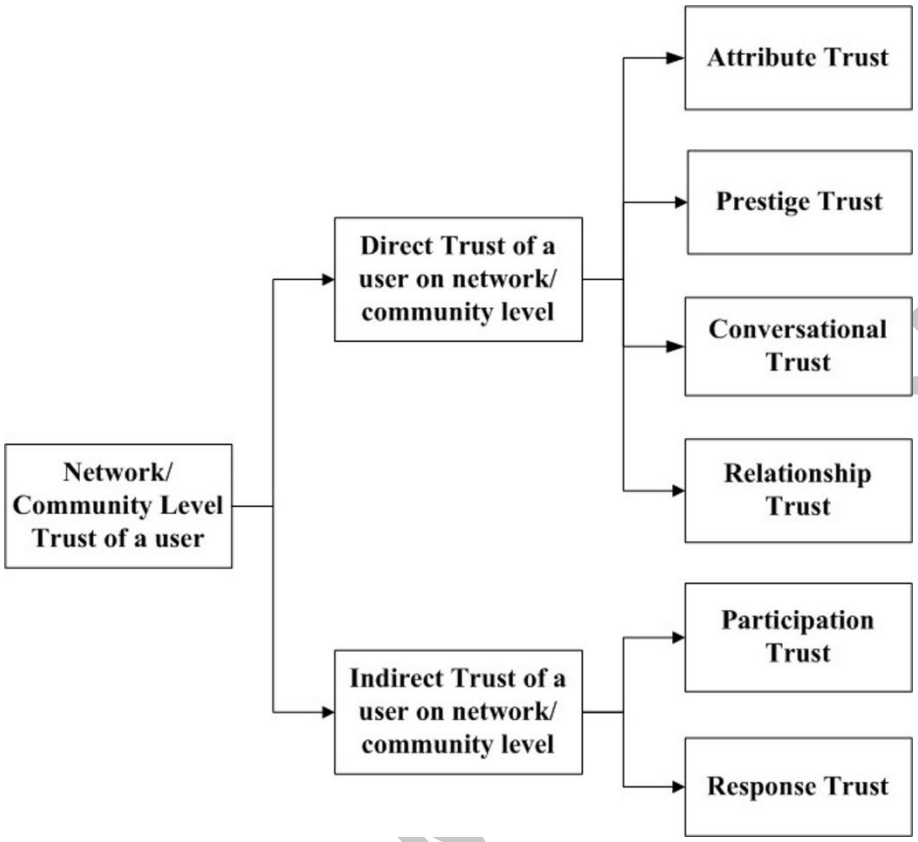


Fig. 3: User Trust Calculation

4.1 Direct Trust Calculation

Direct trust (or local trust) is the belief in a user calculated by direct experience of trustor with trustee (Lewis, 1985). Let $G\ (V, A, AT)$ is a weighted graph, where V represents set of vertices, A represents set of arcs between them and AT represents set of attributes associated with each vertex $\in V$. Each vertex u is having arcs, where $a = (u, v)$ represents an arc from vertex u to v having weight $w(a)$. In *SNTrust* model, *Direct Trust* is composed of four trust values that are calculated for each individual in the network. Calculation of four *Direct Trust* components is described below that include attribute trust, prestige trust, conversational trust and relationships trust.

A. Attribute trust

It is found that people having similar characteristics feel more comfortable and are close to each other (Baek and Kim, 2014). Users mostly trust others having common/similar attributes (Golbeck, 2009), so attributes of users can be used to find their trust. Each user u can have N attributes where $N = |Attributes_u|$.The attribute trust of u from each of his neighbor v is calculated by the Equation 1.

$$AT_{u \to v} = |Attributes_u \cap Attributes_v| / |Attributes_u| \qquad (1)$$

After calculating attribute similarity of node u to each of his neighbor v, the average value of attribute trust for u is calculated by the Equation 2.

$$avg.AT(u) = \forall v \in adj(u) \Sigma\ AT_{u \to v}/n \qquad (2)$$

Where n denotes a total number of neighbor nodes of u.

B. Prestige trust

Interactions between users of OSN can also be called "popularity-based interaction" which can show trustworthiness of a user in his group (S. Nepal, 2011). Degree prestige can be a true representative of it. It tells us in-degree of a node. Ties between actors can include positive or negative values in directed networks. It is not necessary that in-degree of a node shows its degree prestige. It depends on scenario taken for analysis. For example, if a tie has a negative value and it is in-degree of a node then it can't be the representative of a node prestige. Moreover, if a tie represents advice given from one person to another then out-degree of sender node will be used for calculating degree prestige.

High values of direct and positive connections (e.g., likes) to a person will be a true representative of the trust from other persons for him. It can be also called as the "positive view of some actor". In short, we are using degree prestige for trust calculation of an actor which will tell attitude of other actors towards that actor. The degree prestige is calculated using Equation 3.

$$PresT(u) = \begin{cases} 1, & \textbf{\textit{if}}\ Pos.Arcs_{v\to u}(u) > Neg.Arcs_{v\to u}(u) \\ 0, & otherwise \end{cases} \quad (3)$$

Where $Pos.Arcs_{v\to u}(u)$ as mentioned in Equation 4 represents the sum of all positive arcs coming from all neighbors $v \in adj(u)$ towards node u. $Neg.Arcs_{v\to u}(u)$ which is calculated in Equation 5 denotes the total negative arcs coming towards u from his neighbors and n is total in-degree of node u.

$$Pos.Arcs_{v\to u}(u) = \sum_{\forall v \in adj(u)} a(v,u) \ \ where\ w(a) > 0 \quad (4)$$

$$Neg.Arcs_{v\to u}(u) = n - Pos.Arcs_{v\to u}(u) \quad (5)$$

C. Conversational trust

The behavior of a user is the main source of trust information in a SN and includes the type of interaction, the frequency of interaction etc. (Sherchan, 2013). Moreover, the frequency of interaction of a user towards an actor can also contribute to that actor's trust from other users (Volakis, 2011). It is a type of behavioral trust of a user. It tells the length of communication and frequency of communication between two users (Adali et al., 2010). Here, we are only analyzing the frequency of communication between two users. The frequency of communication between two users is directly proportional to the trust between them. The Equation 6 is used to calculate the conversational trust of u from all his neighbors v.

$$CT(u) = \frac{[\forall v \in adj(u) \sum_{v\to u} w(v,u)\ ]}{w(A)} \quad (6)$$

Here $w(v,u)$ is weight of arc that is coming from neighbor v to u. Weight of an arc denotes the frequency of communication between vertices v and u. Greater weight denotes high communications between u and v. The $w(A)$ is the sum of weight the of all arcs in weighted graph $G(V, A, AT)$ as shown in the Equation 7.

$$w(A) = \sum_{\forall (a) \in A} w(a) \quad (7)$$

D. Relationship trust

User-to-user relationships are of great importance when trust is considered. In this paper, these relationships are classified into two categories: Natural relationships and Dynamic relationships. Natural relationships include family relationships and Dynamic relationships are those relationships that can change with the passage of time, for example, "friend" relationship between two users can be changed to "best friend" or to "Not a friend" in the future. Table 1 lists the user relationships, according to categories.

Table 1: User-to-User Relationships Taxonomy

Category	**Relationships**	
Natural relationships	Blood relations, Family relationships (father of, mother of, sister of, grandfather of, wife of, child of, fiancé of, husband of, brother of, uncle of, aunt of).	
Dynamic relationships	Friendships	Close friends, mutual friends, and friends.
	Workplace relationships	Colleagues, Boss, Teammate, Subordinate, Partner, co-worker.
	Geographical Relationships	A neighbor of, of similar hometown, of the similar city, of the similar province, of the similar country, of a similar region.
	Follower relationship	Follower of, followed by.
	Casual relationships	Has met, has the know-how, has seen somewhere.
	Restricted relationship	Never seen, don't want to know, the enemy of.

The following Algorithm 1 is used for relationship trust calculation.

Algorithm1: Relationship Trust calculation algorithm

Input: Node u, and the set V of its neighbors
Output: The array of relationship trust values RT of u with each neighbor $v \in V$
Begin:
For all neighbors $v \in V$ of u **do**
$r \leftarrow extractRelationship\ (u, v)$
if $r \in natural\ relationships$
RT $_{u \to v}$=1
else if $r \in friendships$
if r = = close friends
RT $_{u \to v}$=.95
else if r = = friends
RT $_{u \to v}$=.90
else
RT $_{u \to v}$=.85
else if $r \in workplace\ relationships$
RT $_{u \to v}$=.80
else if $r \in geographical\ relationships$
RT $_{u \to v}$=.60
else if $r \in follow\ relationships$
RT $_{u \to v}$=.40
else if $r \in casual\ relationships$
RT $_{u \to v}$=.20
else
RT $_{u \to v}$=0
end if
save relationship trust value between u and v
end for
End

Now, if a node u has five neighbors then five relationship trust values are obtained using the above algorithm. The Equation 8 is used to find the average value of relationship trust of node u with respect to its neighbors n.

$$avg.RT(u) = \forall v \in adj(u) \sum RT(u, v)/n \quad (8)$$

At the end, the final value of direct trust for a node u is calculated by the Equation 9 where, $\alpha = \beta = \gamma = \delta = \frac{1}{4} = 0.25$.

$$DT(u) = \alpha.avg.AT(u) + \beta.PresT(u) + \gamma.CT(u) + \delta avg.RT(u) \quad (9)$$

If the direct trust on a community level is required, then the Equation 10 can be used. The only difference between direct trust of a user on a network and community level is that the information of the nodes in a community, their connections, and their social relationships within that community, and attributes are used in the latter case.

$$DT(u)_g = \alpha.avg.AT(u)_g + \beta.PresT(u)_g + \gamma.CT(u)_g + \delta avg.RT(u)_g \quad (10)$$

We use a threshold of trust as 0.5. If a user has direct trust value, $(u) \geq 0.5$, then he is considered trusted user on individual level (locally), otherwise untrusted.

4.2 Indirect Trust

Indirect trust of a user is calculated at the global level in a network in contrast to the direct trust calculation at the local level. Indirect trust (Global trust) value of a user is calculated on the basis of user activities and the responses of other users against those activities. The indirect trust consists of participation trust and response trust of each user.

A. Participation trust

Engagement-based interactions between users depict how much a user trust others in the network (S. Nepal, 2011). It can be analyzed by the extent to which a user is involved in sharing posts at network/group level. If a user's total posts in a network/group are less than the average number of posts in that network/group respectively, the user is considered as a non-participative user otherwise participative. Participation trust (PT) of each node is maintained at network/group level and is calculated by using the Equation 11.

$$\mathrm{PT(u)} = \begin{cases} 1 & \text{if } \mathrm{P(u)} \geq \mathrm{AP} \\ 0 & \text{otherwise} \end{cases} \qquad (11)$$

Here, $P\ (u)$ shows the total number of posts of node u in a network/group which is calculated using Equation 12.

$$P(u) = \sum_{i=1}^{n} P_i \qquad (12)$$

Moreover, AP denotes the average number of posts per user on network/group level. If AP is required on the network level then Equation 13 can be used, otherwise Equation 14 can be used to determine the average number of posts per user on group level.

$$AP = \frac{\sum_{i=1}^{n} P_N}{n_N} \qquad (13)$$

$$AP = \frac{\sum_{i=1}^{n} P_g}{n_g} \qquad (14)$$

Here, $\sum_{i=1}^{n} P_N$ is the total number of posts in a network and n_N is total number of users in a network. Whereas $\sum_{i=1}^{n} P_g$ is the total number of posts in a group and n_g is total number of members in that group.

B. Response trust:

Trust of a user can be calculated by examining his reputation that comes from other users interacting with her (Wang and Liu, 2007). Conversational trust of each user is calculated at network/group level. If a user shares post in a network/group, his indirect trust calculation is also possible using the reactions received on her posts by other users/group members. For response trust, the number of positive and negative responses (reactions) on a user's post is considered. Responses are categorized and are given weights as mentioned in Table 2.

Table 2: Categories of user responses against a post

Response category	Possible Responses(acts)	Response weights (w_{act})
Positive response	Share	1
	Comment	.75
	Love	.75
	Wow	.50
	Like	.25
	HaHa	.25
Negative response	Angry	-1
	Dislike	-0.50

Response trust (ResT) of a user post up is calculated as the weighted sum of all reactions (positive or negative) on that post divided by the total number of reactions on that post as shown in the Equation 15. Similarly, the response trust of a user on a community level can only be calculated by considering the actions against the posts of that user in that community from other community members instead of the complete network.

$$ResT(u) = \forall v \in Network \left(\frac{\sum_{act \in ACT} \left(act\left(v_{up}\right) * w_{act} \right)}{\sum_{\forall act \in ACT} up} \right) \qquad (15)$$

Where $act\left(v_{up}\right)$ denotes the action of user v against the post up of user u and w_{act} denotes the weight assigned to te action (act) in the action set ACT. The average value of response trust for a user u is calculated for all n posts of u in a network/group as shown in the Equation 16.

$$avg.ResT(u) = \frac{\sum_{i=1}^{n} ResT_i(u)}{\sum_{i=1}^{n} up} \qquad (16)$$

Where $ResT_i(u)$ is response trust of user u against his one post. The sum of all posts of user u is divided by the total number of posts of u in a network/group. If $avg.ResT(u) \geq 0.5$ then u is receiving good response trust otherwise not. The greater value of response trust indicates that posts of the user u are liked and considered worthy by other members of his group. Indirect Trust (IT) of each user u on a network/community level is calculated by combining participation trust (PT) and response trust (ResT). The Equation 17 is used for calculation of a user's indirect trust, where $\omega = \mu = 0.5$.

$$IT(u) = \omega PT(u) + \mu.avg.ResT(u) \qquad (17)$$

If $IT(u) \geq 0.5$, the use u is considered trusted user at nework/group level, otherwise not. Greater user trust indicates that the user is an active participant of the network/group and user's posts are also appreciated by others. If a user is not participating in the network/group and getting no responses from his group members, the user is considered as non-active member.

4.3 Community/Group level trust

Group level trust (GT) of each user u is calculated by combining its direct trust (DT) and indirect trust (IT) on a community level. The Equation 18 is used for calculation of user trust on group level, where $a = b = 0.5$.

$$GT(u) = aDT(u)_g + bIT(u)_g \qquad (18)$$

If $GT(u) \geq 0.5$, the use u is considered trusted user at group level, otherwise not. If a user is the member of more than one group then her group trust values can be different with respect to groups. The average value of a user u based on the participation in many groups is calculated by the Equation 19.

$$avgGT(u) = \frac{\sum_{\forall g \in G} GT(u)}{G} \qquad (19)$$

Here, G represents the total number of groups in which a node u is involved and $GT(u)$ shows the group level trust of user u in any group g in the network.

4.4 Network level trust

After the calculation of dyad trust (direct trust) and indirect trust of each user on the network level; both trust values are used to find network level trust of the user. Network level trust of each user is calculated by using Equation 20 where $c = d = 0.5$.

$$NT(u) = c.DT(u)_N + d.IT(u)_N \qquad (20)$$

If $NT(u) \geq 0.5$, then a user u is considered as a trusted user on the network level, otherwise non-trusted.

It is worth mentioning that the weights given to different trust values in Equation 9, Equation 17, Equation 18, and Equation 20 can be determined based on the respective scenario. If all mentioned types of trust are equally important for a user, then equal fix weightage to each of trust values seems reasonable. However, depending upon the nature of the environment in which the trust is being calculated and also the needs of a user, the weights given to different types of trust may vary i.e. some trust values can be given more weights than others. For example, if a user A wants to be a friend of a user B, he may or may not consider the user B's attribute similarity important for him. Likewise, if a user A is a brother of user B, then he may only consider that relationship important to place his trust on user B by ignoring other types of trust.

5. Results and Discussion

SNTrust model is implemented and standard datasets are used for results evaluation. SNTrust model results are compared with influence results obtained from influential node identification methods such as k-core, Closeness centrality, Page Rank, and Eigenvector. Louvain algorithm is used to perform community detection in the network. The results related to preset objectives are mentioned in section 4.1 and indirect trust calculation is performed in section 4.2.

5.1 Analysis of Trust and Influence

In this subsection, details of aforesaid datasets are given and the results are discussed to attain the targeted research objectives.

5.1.1 Consulting company dataset

A. Dataset description[1]

The dataset consists of two related networks about 46 employees of a consulting company. In the first network, there are 879 ties which show the frequency of advice requests between 0 and 5 (0: I Do Not Know This Person; 1: Never; 2: Seldom; 3: Sometimes; 4: Often; and 5: Very Often) in response to the question "Please indicate how often you have turned to this person for information or advice on work-related topics in the past three months". In the second network, there are 803 ties which show weights between 0 and 5 (0: I Do Not Know This Person; 1: Strongly Disagree; 2: Disagree; 3: Neutral; 4: Agree; and 5: Strongly Agree) in response to the question "Please show how strongly you agree or disagree with: The person in the given list has expertise in areas that are important in the kind of work I do". Furthermore, Table 3 shows the attributes of the employee and their values.

Table 3: Attributes of employees in Consulting company dataset

Attribute Name	Attribute Value
Organization Level	1. Research Assistant, 2. Junior Consultant, 3. Senior Consultant, 4. Managing Consultant. 5. Partner
Gender	1. Male, 2. Female
Region	1. Europe, 2. USA
Location	1. Boston, 2. London, 3. Paris, 4. Rome, 5. Madrid, 6. Oslo, 7. Copenhagen

B. Results and Evaluation

At first, direct trust (local trust) values for all employees in the network are calculated using SNTrust model and are compared with results of many influential node identification techniques. The Fig. 4 shows the results of direct trust for all nodes and categorizes them into trusted and non-trusted nodes which resulted in 33 and 13 respectively. It can be clearly seen that trusted nodes are densely connected than non-trusted nodes, which show that trusted people are more likely to communicate with each other and vice versa. Moreover, non-trusted nodes in the network highly communicate with trusted nodes than non-trusted nodes.

[1] https://toreopsahl.com/datasets/#Cross_Parker

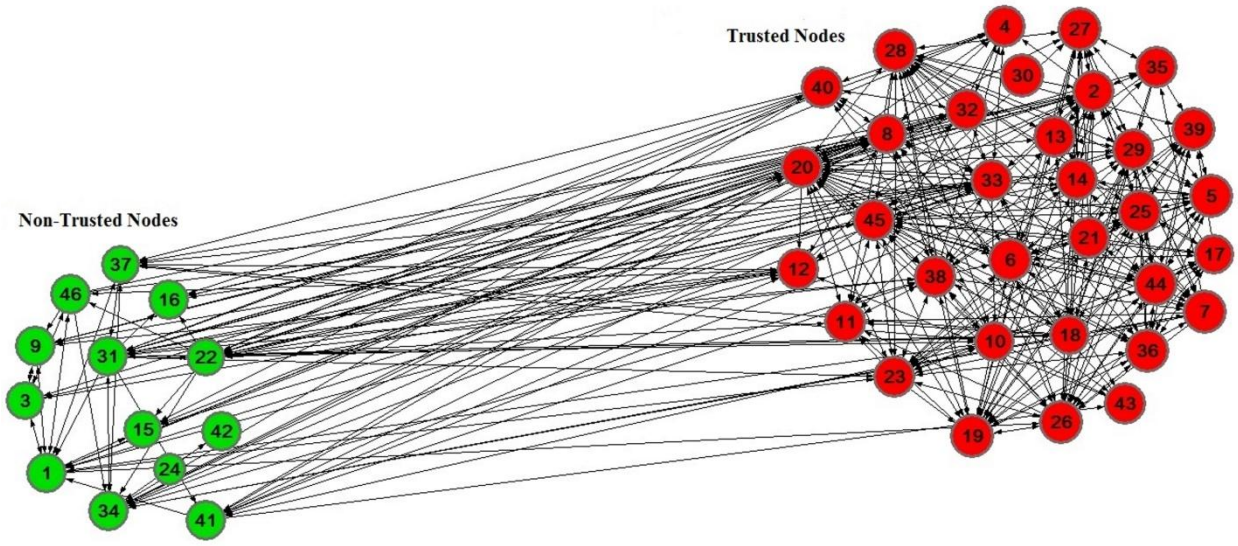


Fig. 4: Trusted/ Non-Trusted nodes in the Network

We have explored further to obtain results for focused research objectives. The outcomes are as follows.

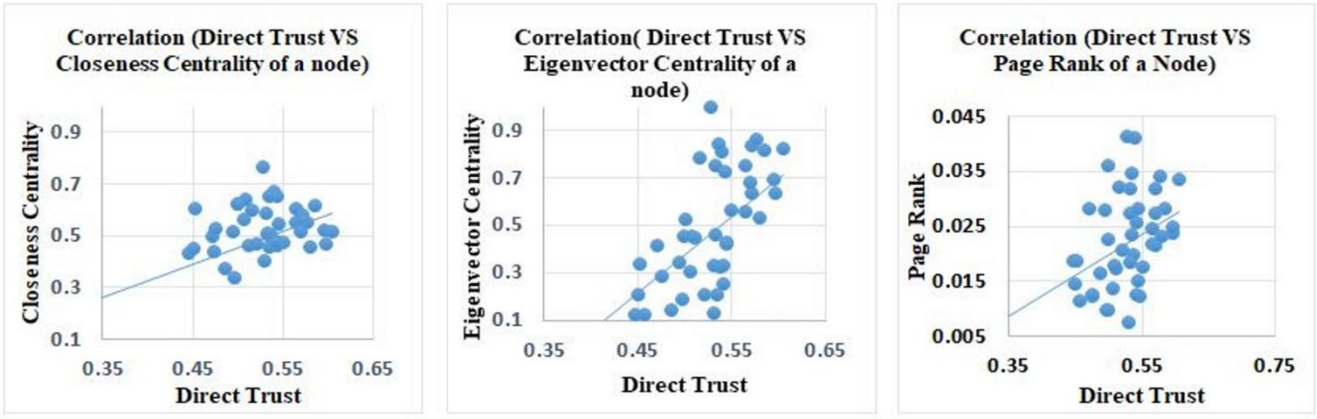
The results from the SN Trust model and influence are compared to find their correlation to achieve our second research objective. We calculated the influence of a node in the network using k-Core, Closeness centrality, Eigenvector centrality and Page Rank. The closeness centrality is the average distance of a node from all other nodes. The eigenvector centrality of a node emphasizes the fact that the friends of influential nodes are also influential. The PageRank based centrality measure is used to determine centrality scores in terms of user's connectivity in the weighted activity graph.

Table 4 shows statistical analysis of the direct trust method with aforementioned influence evaluation techniques. It is found that there is a positive correlation between direct trust and influence. The strength of the relationship is weak in the case of Closeness and PageRank while there is a moderate linear relationship between direct trust and eigenvector. Significance level, also known as p-value, shows that two variables are linearly related if p-value <.05. Here, results show that this condition is true for all cases. It means that the correlation between both approaches for finding node's importance is statistically significant. In other words, it is more likely that trusted nodes are closer (Closeness centrality) to other nodes in the network, have a higher ranking (PageRank) in the network, and have more global prominence (higher eigenvector centrality).

Table 4: Correlation and p-value between Direct Trust and Influential Nodes Detection methods

On Network Level		**Closeness**	**PageRank**	**Eigenvector**
Direct Trust	Pearson Correlation Coefficient (r)	0.375*	0.409**	0.601**
	Significance of Coorelation (p-value)	.010	.005	.000
	Correlation Analysis	*. Correlation is significant at the 0.05 level (2-tailed).	**. Correlation is significant at the 0.01 level (2-tailed).	

The following Fig. 5 shows the linear positive correlation between direct trust of nodes and their measures of influence.



Weak uphill linear positive relationship **(a)**

Moderate uphill linear positive relationship **(b)**

Weak uphill linear positive relationship **(c)**

Fig. 5: Pearson's Correlation Analysis of results

We have also investigated a k-core method for finding the most important nodes in the network. A K-core of a network is the group of nodes where all nodes have at least K degree. It represents the most cohesive part of the network. During analysis, it is found that highest k-core present in the network is 19 having size 14. It means that there are 14 nodes in the maximal dense group in the network and each node of this group is directly connected to other 19 nodes. The Fig. 6 shows 19-core group found in the network. It is worth mentioning that all of the nodes that are part of this 19-core are also indicated as trusted nodes by SNTrust model. It shows that the group of most influential nodes in the network is also trusted. Further, the most trusted node in the network (Node 6) is also the part of 19-core. It can be said that the highest k-core nodes in the network are also trusted nodes.

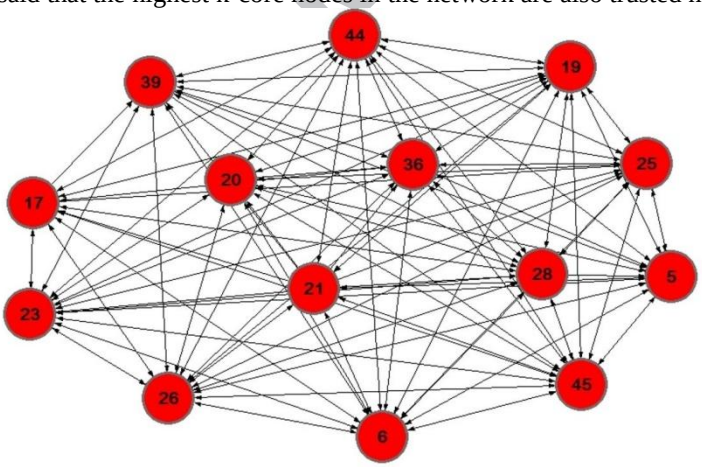


Fig. 6: Most influential nodes using K-core Method

Some authors argue that studying complete network for the influential node detection is time-consuming and has high computational costs (Trusov et al., 2010). This overhead can be minimized by detecting influential nodes in communities as compared to the whole network (Han et al., 2017; Zeng et al., 2016). To find whether highly trusted nodes of the network are also trusted in their own communities, SNTrust model is also applied at the community level. *Louvain* algorithm is used to find communities in the network (Blondel et al., 2008). It uses *modularity* to divide a network into high-quality partitions. The quality of partitions indicates that there are more intra-community connections than the inter-community connections.

It is found that there are five communities in the network as shown in the Fig. 7. The color of nodes indicates their particular communities. The red, green, yellow, blue and pink nodes represent five different communities in the network.

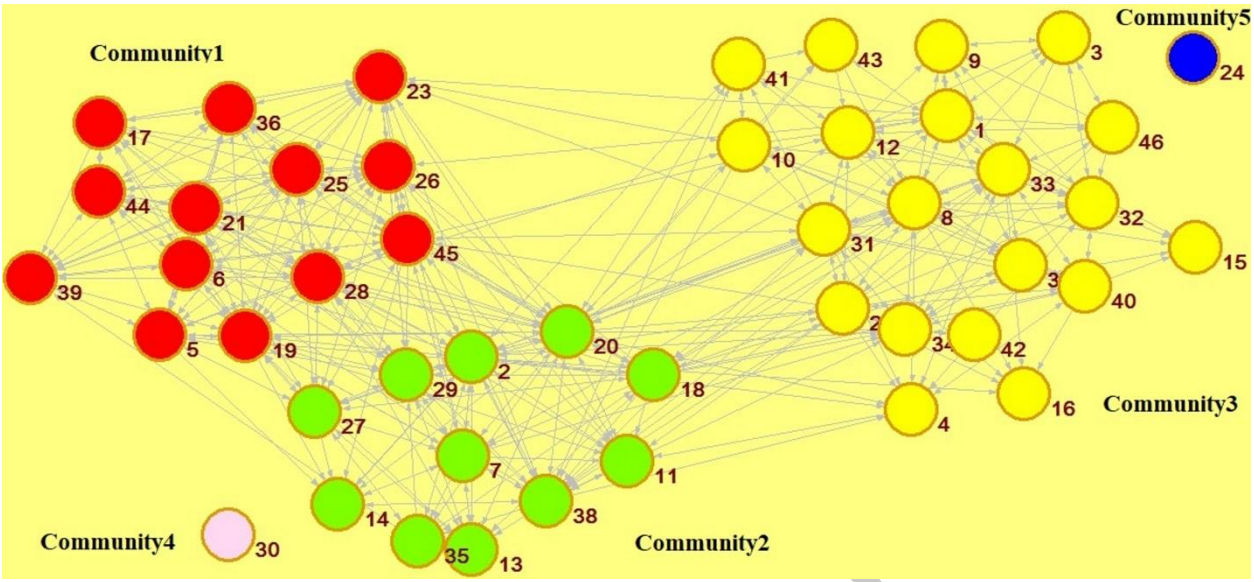


Fig. 7: Communities found in the Consulting Company Network

Direct trust of each node in its community is calculated. It is found that there are 41 nodes having $direct\ trust\ value >= 0.5$ (threshold). By comparing trust of nodes at the community level and network level, it is seen that there are more trusted nodes in their own communities as compared to the whole network. It indicates that employees of consulting company tend to communicate more in their own communities as compared to communication between communities. The Fig. 8 shows that direct trust of nodes within their communities is high as compared to their trust in the whole network.

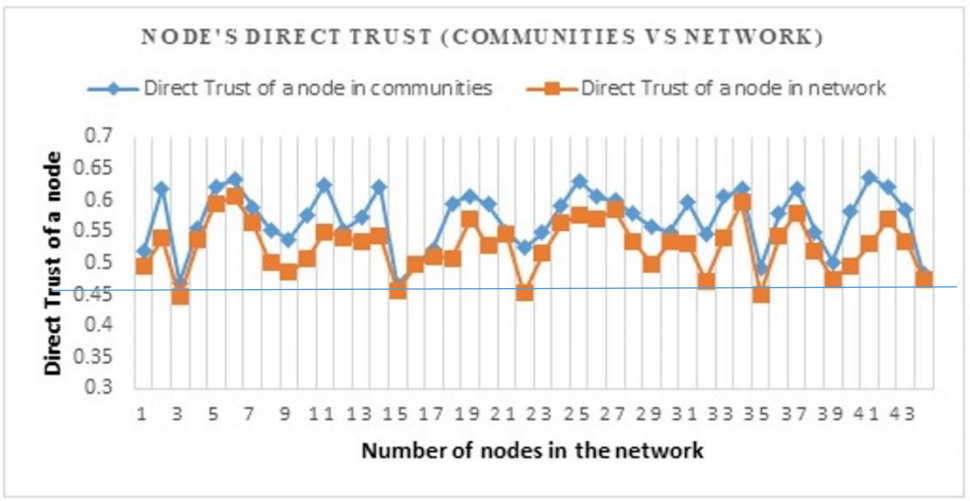


Fig. 8: Comparison of Direct Trust in communities and network

As we have analyzed the correlation between trust of a node in its community and in the network, it is found that Pearson's correlation produces 0.83 value which show high correlation in both variables. It indicates that the nodes having high direct trust in the network are also highly trusted in their communities. The relationship between both variables is statistically significant, $r(44)=0.83$, $p<0.001$. The Fig. 9 shows the strong positive linear relationship between the trust of a node in its network and the trust of a node in the community.

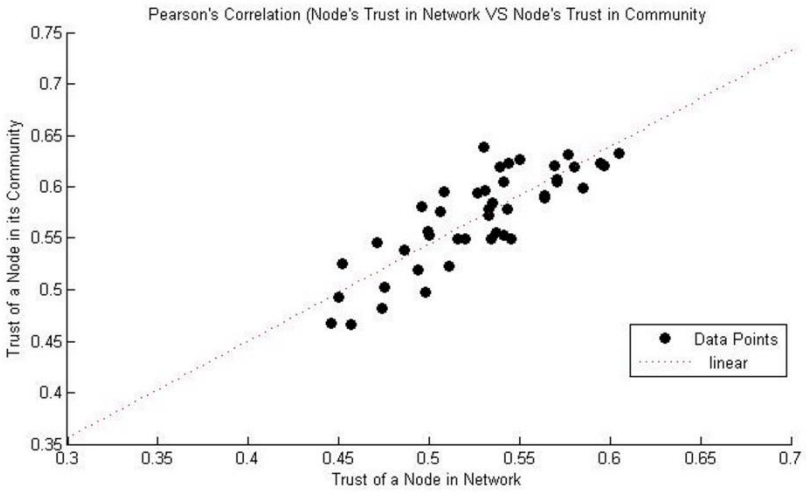
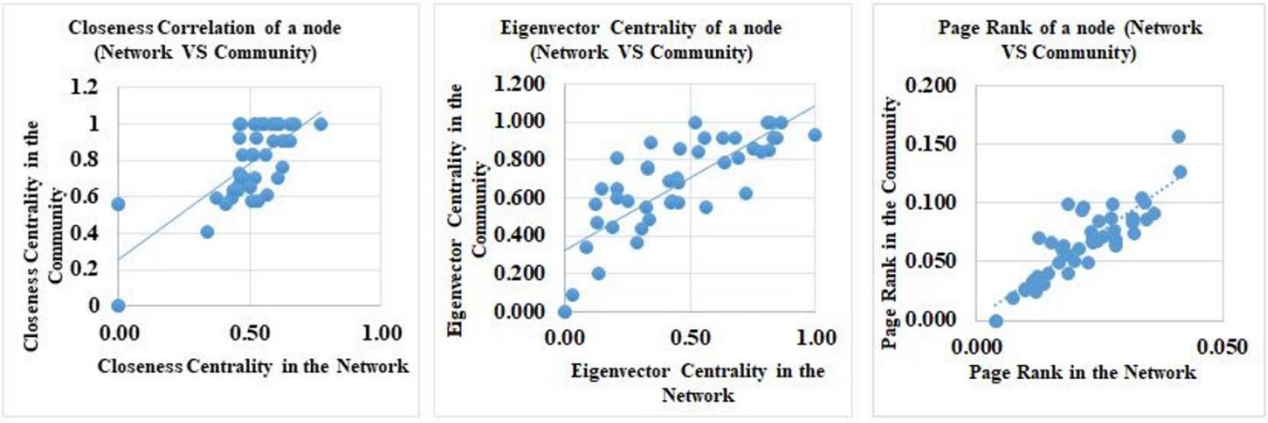


Fig. 9: Pearson Correlation Analysis of node trust in community and network

To achieve our fourth objective, Pearson correlation is calculated between the influence value of a node in the whole network and in its community. It is worth mentioning here that there is a strong positive linear relationship between the influence of a node in the network and community. It indicates that the nodes having a high influence in the network are also highly influential in the communities. The relationship between both variables is statistically significant, where it is r (44) = 0.76, p<0.001 in case of closeness centrality, r (44)=0.79, p<0.001 in case of eigenvector centrality, and r(44)=0.86, p<0.001 in case of Page Rank. The results are shown in the Fig. 10.



(a) **(b)** **(c)**

Fig. 10: Pearson correlation analysis of node influence in communities and network

C. Analysis of Local Structure of Network against Local Trust

To analyze the local structure (communities) in the network, average direct trust between members of these communities is calculated. Different SN measures such as the average closeness of nodes, average eigenvector centrality of nodes, number of triads and density are compared with the average direct trust of each community. Table 5 shows some interesting findings. Results show that average direct trust in a community is directly proportional to the average closeness as well as average eigenvector in that community. Moreover, there are more triads in a community having high average direct trust among its members. Triad represents that friend of a friend are also friends also which means the connection between the three nodes is high. The density of a community is also directly proportional to average direct trust. High density means that there are more dyadic connections among nodes. In short, it can be said that if the members of a community have high direct trust, they are much closer, transitive and connected with each other.

Table 5: Analysis of Local Structure of Consulting Company Network

Communities	Average Direct	Average	Average	Number of	Density of

	Trust in community	closeness in community	Eigenvector in community	triads in community	community
Community1	0.565	0.988	0.863	180 triads	0.85
Community2	0.541	0.862	0.747	39 triads	0.62
Community3	0.498	0.652	0.572	38 triads	0.34

The Fig. 11 represents nodes in their communities with respect to their closeness centrality. The color of nodes depicts the community to which they belong and size as well as label represents the value of closeness centrality. It is observed that most of the nodes in *Community1* have a high closeness centrality and have more connections among them. Whereas nodes in *Community3* vary in sizes with respect to their closeness centralities and there are sparse connections in this community. Most of the network edges are either inside *Community1* or coming to/from *Community1*. The centrality values show that nodes in the *Community1* are very close to each other. It is again mentionable here that average trust in Community1 is higher than other communities.

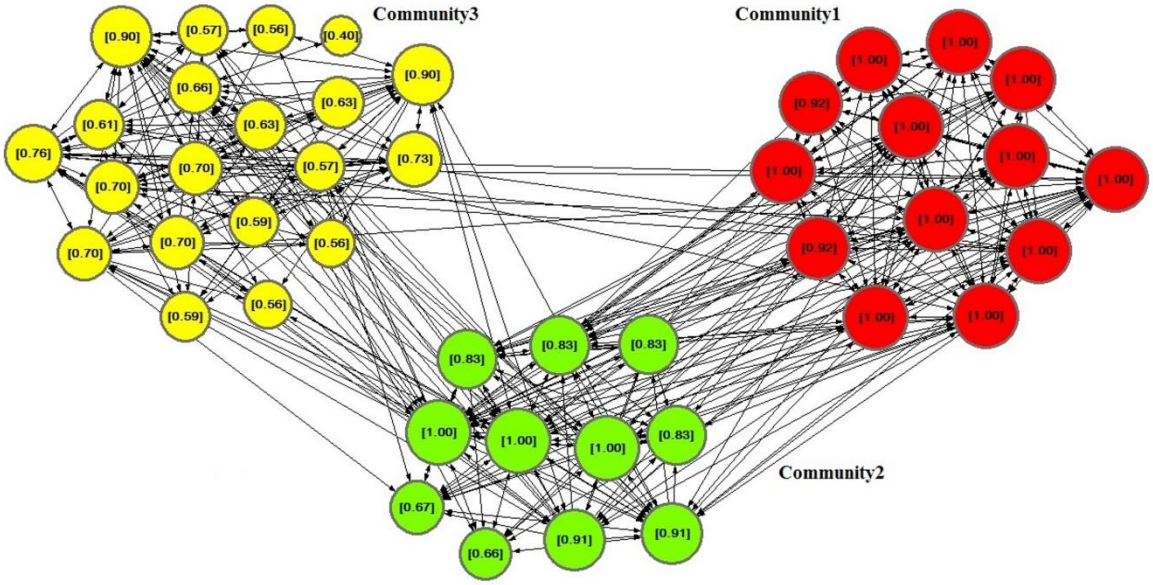


Fig. 11: Closeness centralities at the community level

As we have seen in the results (related to second research objective) that eigenvector centrality of a node has a linear relationship with its direct trust. The Fig. 12 shows that nodes in the *Community1* have more trust on each other and it can be noticed that most of the nodes have high eigenvector centrality (large size). Moreover, nodes with small sizes (less prominence) are the part of *Community3* having least trust among its members.

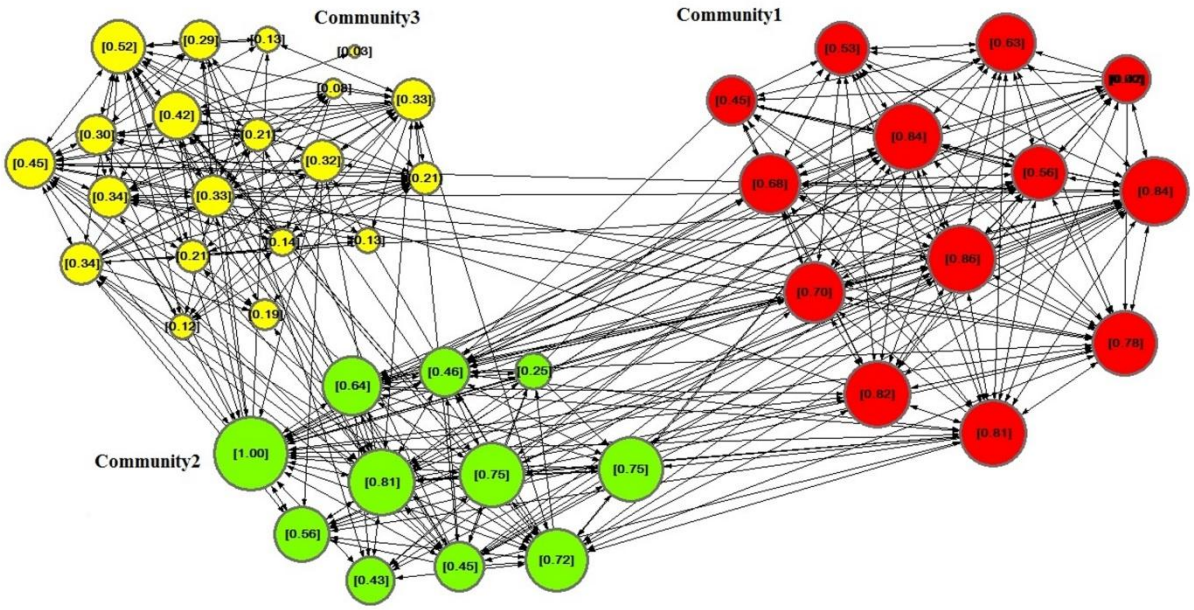


Fig. 12: Eigenvector centralities at the community level

D. Demographic patterns found in the Network

We have also analyzed the network to find patterns in its different communities. Here, some interesting results are found. Table 6 depicts the attributes of each employee with respect to his community. It is clear from the table that in *Community1*, all people belong to Boston, USA, most of them are male and playing consultant role in organizations. It is quite interesting as we have seen earlier that average group trust and average centrality values are high in *Community1* as well. It can be said that the persons having a same region and organizational roles are more likely to trust each other and they are closer as well. On the other hand, members of *Community3* belong to different countries of Europe having different organizational roles. It is stated earlier that the members of *Community3* have minimum average direct trust and average centrality values. It indicates that the persons with different geographical locations are less likely to trust each other.

Table 6: Frequency Table of demographic features w.r.t. communities

Attribute Name	Attribute value	Network	Community1	Community2	Community3
Gender	Male	35	10	8	16
	Female	11	3	3	4
Region	Europe	20	---	---	20
	USA	26	13	11	---
Organization Level	Research Assistant	6	1	---	3
	Junior Consultant	9	1	4	4
	Senior Consultant	10	4	3	3
	Managing Consultant	17	6	3	8
	Partner	4	1	1	2
Location	Boston	26	13	11	---
	London	1	---	---	1
	Paris	9	---	---	9
	Rome	2	---	---	2

Madrid	2	---	---	2
Oslo	3	---	---	3
Copenhagen	3	---	---	3

5.1.2 Freeman EIES Dataset

A. Dataset Description[2]

This dataset contains results of a longitudinal study related to researchers. The personal relationship information of each researcher is recorded at the beginning (at time t1) as well as at the end of the study (at time t2). There are three networks in this dataset. First network contains the 650 relationship ties between two individuals at time t1, which are weighted from 0 to 4 by researchers themselves (0 represents a person unknown to the researcher, 1 represents a person the researcher has heard of but never met, 2 represents a person the researcher has met, 3 represents friendship, and 4 represents close friendship). The second network contains 749 relationship ties between researchers at the end of the study. The third network contains the number of electronic messages exchanged between researchers during the entire period of study. The attribute information consists of the names of researchers, their number of citations, and main disciplinary affiliation. Discipline code used in the dataset is from 1 to 4 (1 represents sociology, 2 represents anthropology, 3 represents mathematics or statistics, and 4 represents others).

B. Results and evaluation

The direct trust and influence of each researcher are calculated. The direct trust of a node at time t1 (at the beginning of the study) is calculated by using attribute trust and relationship trust (personal relationships at time t1) of a node. The direct trust of a node at time t2 (at the end of the study) is calculated by using attributes, personal relationships at time t2, and frequency of communication. It is found that there are 18 trusted nodes at time t1 and 24 trusted nodes at time t2 in the network. The Fig. 13 shows the trusted and non-trusted nodes at the start of the study.

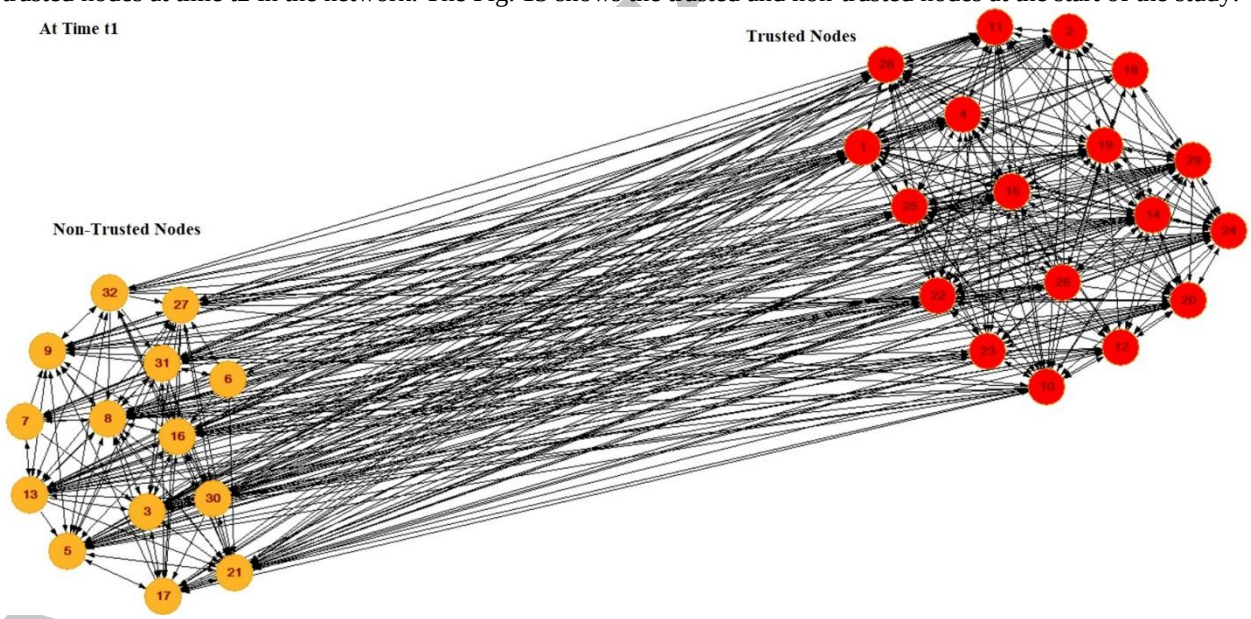


Fig. 13: Trusted/Non-Trusted nodes at the beginning of study

The Fig. 14 shows the trusted and non-trusted nodes in the network at the end of the study.

[2] https://sites.google.com/site/ucinetsoftware/datasets/freemanseiesdata

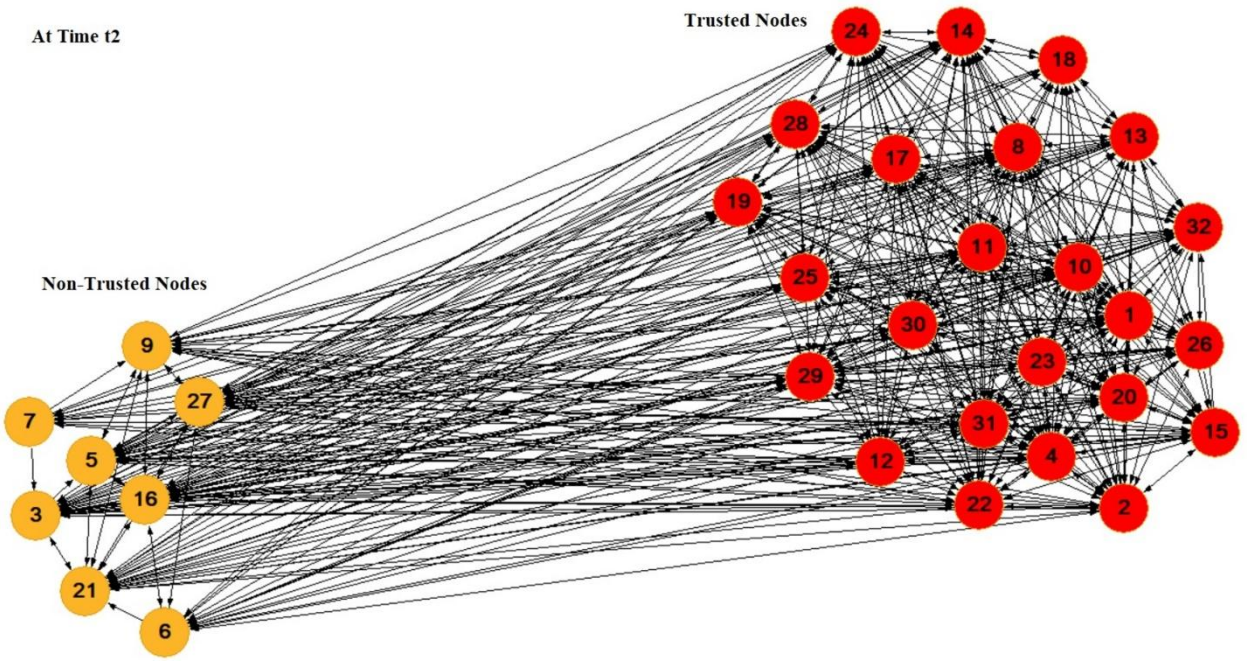
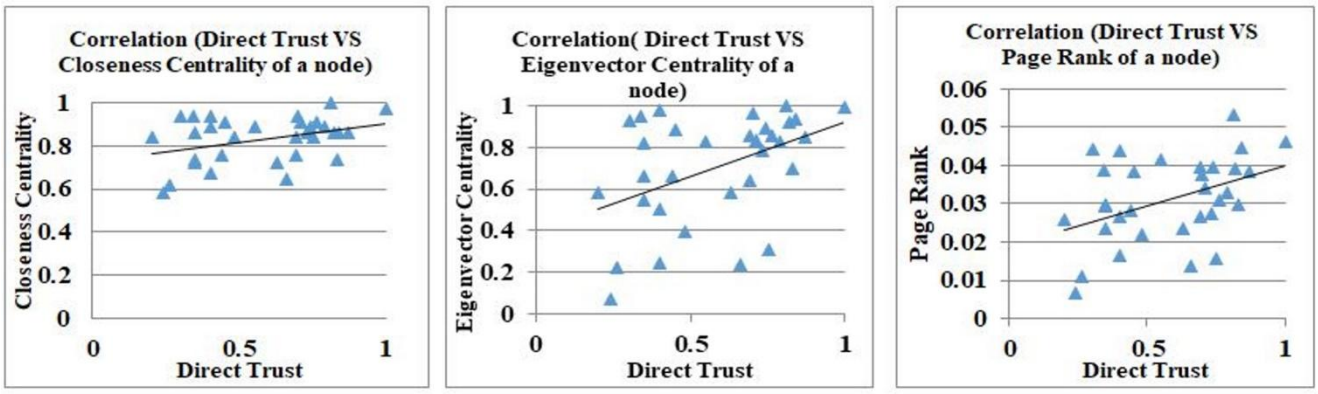


Fig. 14: Trusted/Non-Trusted nodes at the end of study

It can be clearly seen that at time t2, after electronic messages passed between researchers, trust is increased among them. There is less number of ties among non-trusted nodes as compared to trusted ones.

After the calculation of direct trust and influence of researchers, correlation and level of significance are measured between them. At time t1, there is a weak linear positive relationship between direct trust and influence measures (Closeness, PageRank, and Eigenvector) as shown in the Fig. 15. It is also seen that correlation between trust and influence are statistically significant (linearly related) having $p - value < .05$.



(a) **(b)** **(c)**

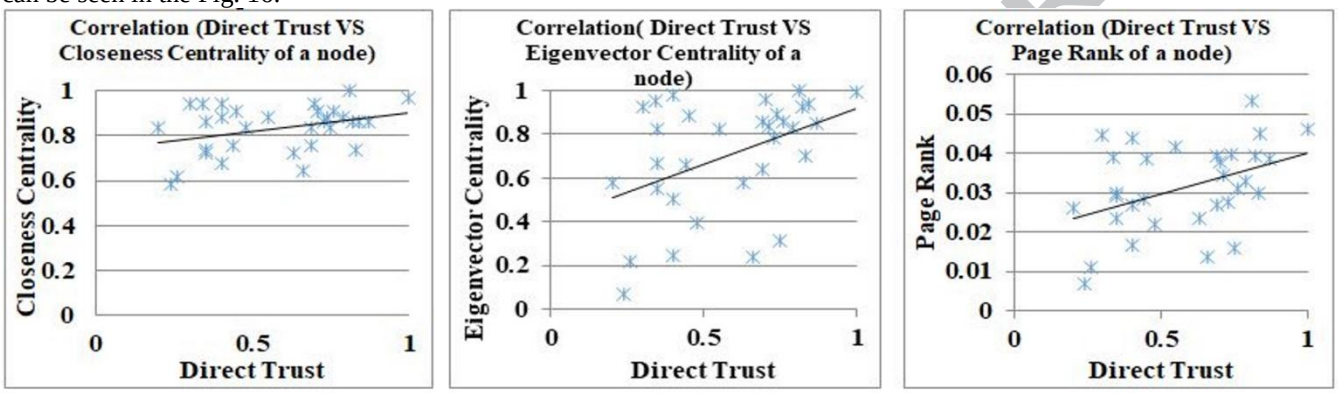
Fig. 15: Pearson's Correlation Analysis: Weak uphill linear positive relationship (at time t1)

It is found that there is a moderate linear positive relationship ($r >= 0.5$) between direct trust of a researcher and his/her influence at time *t2* as shown in the Table 7. It indicates that communication between nodes can lead to stronger association between direct trust and their influence. It is found that there is a significant relationship between trust and influence of a node ($p - value < .001$). In short, results show that if the researchers are trusted, it is more likely that they are also influential.

Table 7: Correlation and p-value between Direct Trust and Influence on network level

At time t2		Closeness	PageRank	Eigenvector
Direct Trust	Pearson Correlation Coefficient (r)	0.5**	0.516**	0.5**
	Significance of Coorelation (p-value)	.0008	.002	.004
	Correlation Analysis	**. Correlation is significant at the 0.01 level (2-tailed).		

At time *t2*, the positive relationship at network level between direct trust and different influence detection methods can be seen in the Fig. 16.



(a) **(b)** **(c)**

Fig. 16: Pearson's Correlation Analysis: Moderate uphill linear positive relationship (at time t2)

We have used k-core to find the influential nodes in personal relationship networks of researchers at time t1 and time t2. The results show that at the beginning of the longitudinal study there is a 17-core group of size 23 (each of the 23 researchers has relationships with at least seventeen others). But, it is important to mention that at the end of the study a 19-core group of size 28 is found. It indicates that the relationships between participants are likely to increase/strengthen after interactions among them. The results of k-core can be seen from following Fig. 17. Moreover, the 23 nodes out of 28 nodes present in the maximal dense group (19-core) of the network are indicated as trusted nodes by SNTrust model. It can be said that trusted nodes can be found in the densest part of the network.

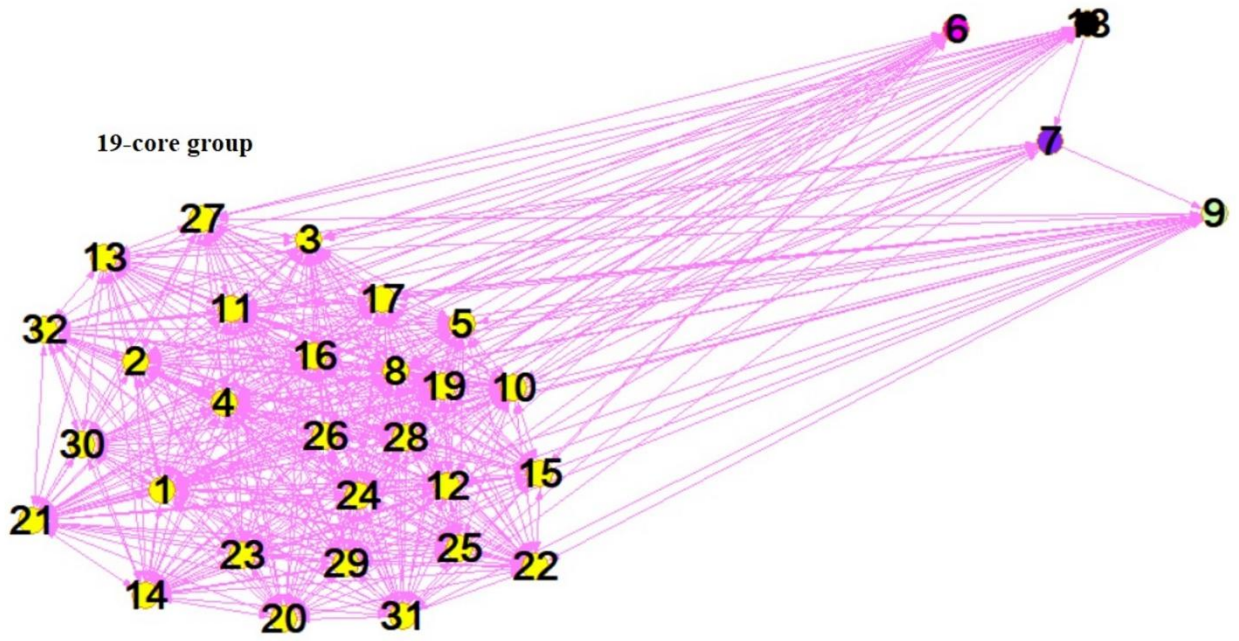


Fig. 17: K-core results at time t2

There are three communities found in the personal relationship network at the beginning of the study. The *Community1*, *Community2*, and *Community3* consist of 19 nodes, 8 nodes, and 5 nodes respectively. At the end of the study, two communities of researchers are found where *Community1* has 18 nodes and *Community2* has 14 nodes. These results show that the number of community members increased after communication among researchers and relationship bond between individuals in the network has become stronger. To achieve our third objective, the direct trust of all nodes at the network level and community level is determined. The Pearson's correlation between direct trust at the network level and direct trust at community level at time t1 are 0.59 and at time t2 is 0.7. It shows that there is a correlation between direct trust of a node on both levels and correlation increased at time t2. It indicates that the nodes having high direct trust in the network are also highly trusted in their communities. The relationship between both variables is found statistically significant at both times by $p - val <$ .001. The Fig. 18 shows the two communities in the network at the end of Freeman study.

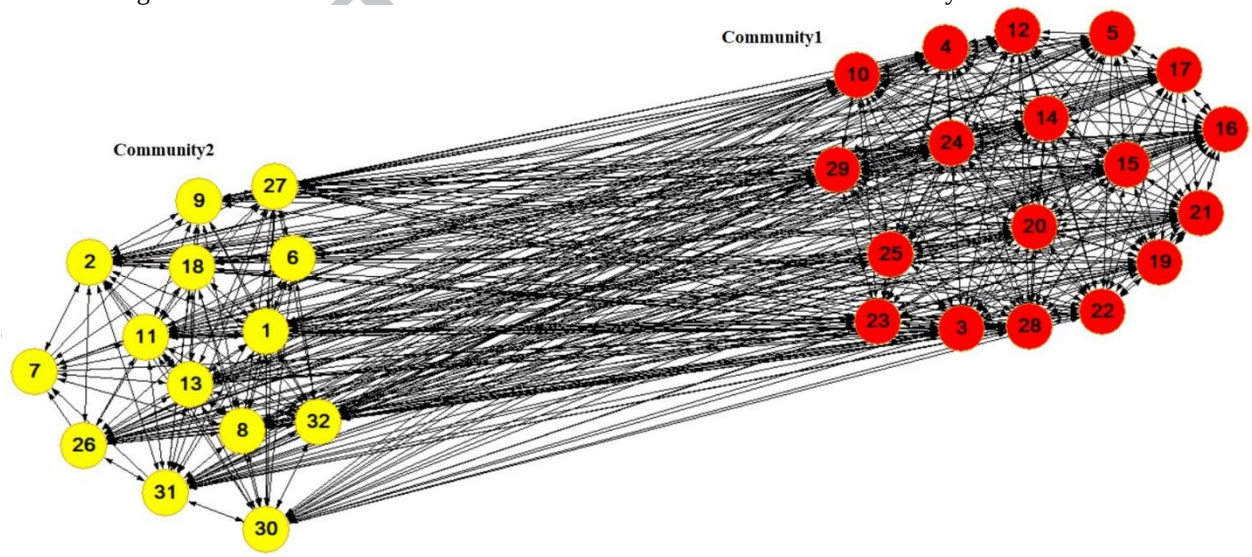


Fig. 18: Two communities found at time t2

The results of influence analysis of a node at the network level and community level can be seen in Table 8. A strong positive linear relationship is found between both variables (influence at the network level and community level). Also, this relationship is statistically significant. It highlights that the influential nodes of the network are also influential in their communities.

Table 8: Pearson's correlation analysis of node influence (Network VS Communities)

Network level VS Community level	Pearson's correlation (r)	P-value
Closeness of a node	0.766	0.000<.001
Page Rank	0.712	
Eigenvector centrality of a node	0.877	

C. Local structure analysis

The results of the local structure analysis of the network are determined by using communities found in the network. As shown in Table 9, it is found that the members of the community, having more direct trust among them are more influential and there are more dyads (connections) and triads (transitivity) in these communities which results in increased density.

Table 9: Analysis of Local Structure of Freeman EIES network

Communities	Average Direct Trust in community	Average closeness in community	Average Eigenvector in community	Number of Triads in community	Density of community
Community1	0.86	0.982	0.926	533	0.86
Community2	0.83	0.894	0.840	169	0.74

We have also analyzed the attribute similarity of the communities found in the network at time t2. The Fig. 19 can show the nodes with their disciplinary affiliations. As shown in Table 9, the *Community1* has more average direct trust and average closeness as compared to *Community2*. It can be clearly seen in the Fig. 18 network that the *Community1* has thirteen researchers having *Anthropology*, three having *Psychology*, and two having *Statistics discipline*. While in *Community2* there is no dominant discipline among researchers. The majority of *Community1* members (72%) have discipline similarity. These results indicate that the people having similar interests tend to trust each other and are closer too. These results are consistent with the study (Golbeck, 2009). Moreover, the density in *Community1* and sparseness in *Community2* is visible.

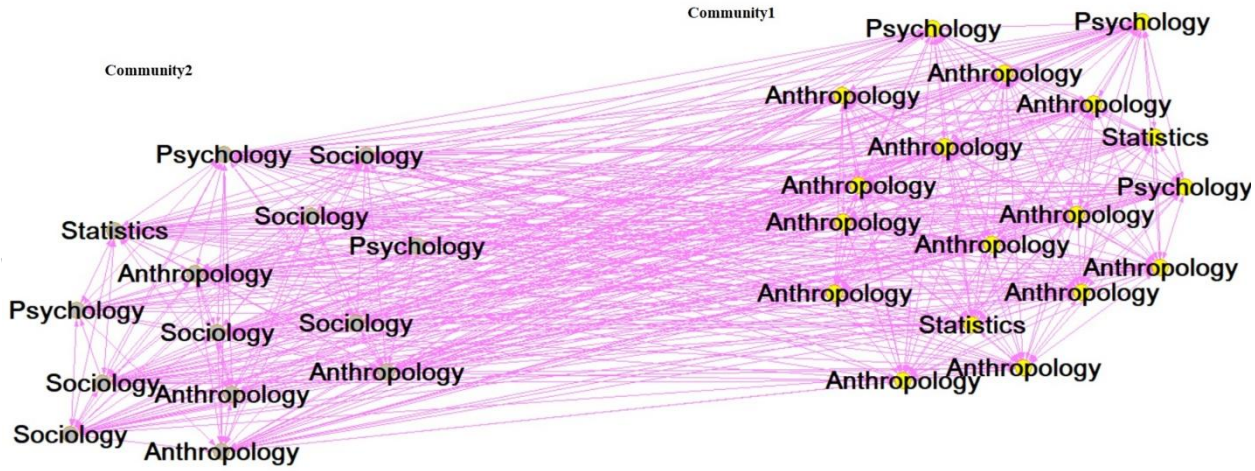


Fig. 19: Discipline similarity in the network communities

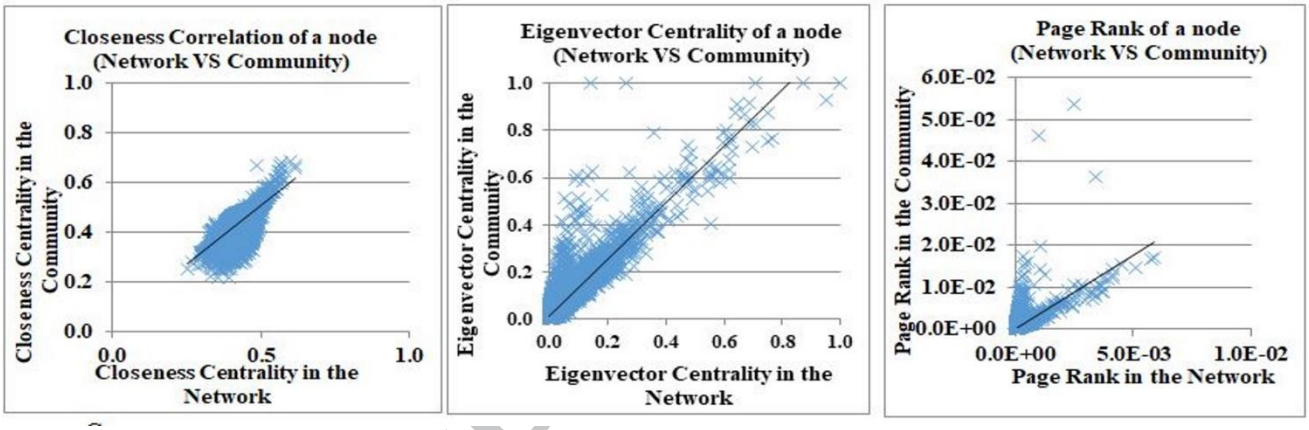
5.1.3 Blogcatalog Dataset

A. Dataset Description[3]

This dataset relates to the bloggers and is collected from a well-known social blog directory named Blogcatalog. It contains the information about 333,983 friend relationships among 10,312 bloggers. It also provides blogger-group membership information about bloggers and there are 39 groups are mentioned in it.

B. Results and evaluation

This dataset has no information about the attributes, frequency of information between bloggers and different kind of relationships between bloggers. Because of this fact, we have used this dataset to only focus on the aforementioned fourth objective. We have calculated the influence of bloggers by using aforementioned algorithms on the network and community level. There are five communities in the blogger's network as a result of Louvain algorithm. Results indicate that strong positive linear correlation is found between the influence of a node on a community level and network level. The Pearson's correlation value (r) is 0.81 in case of closeness centrality; it is 0.91 in case of eigenvector centrality, and 0.7 in case of Page Rank. Additionally, the relationship between both variables is also statistically significant (p-value<0.001 in all three cases). It can be said that the influential bloggers on the network are more likely to be influential on a community level. The following Fig. 20 shows the linear relationship between both influence variables.



(a) **(b)** **(c)**

Fig. 20: Pearson's correlation analysis of a blogger's influence

5.2 Analysis of Indirect Trust based on user activities

The indirect trust of a user (mentioned in Equation 15) can be individually used to find the trusted nodes in a group if only user activity data are provided. The analysis of group level trust is performed on the SN dataset[4] available online. It contains data about three Facebook public groups named Unofficial Cheltenham Township (Group1), Elkins Park Happenings (Group2), and Free Speech Zone (Group3). This dataset consists of data about each of these three groups in terms of group level posts, comments on that post, likes, and responses to posts and the information of group members. There are 1287 members in Group1, 2155 members in Group2, and 245 members in Group3. We have some important observations when the indirect trust calculation was performed on the selected dataset. It is found that 33.95%, 38.37%, and 25.71% members are trusted in Group 1, Group2 and Group3 respectively. It indicates that the members of Group2 are more active in sharing posts and giving a response to other users against those posts. Likewise, there are 66.04% non-trusted members in Group1, 61.62% in Group2, and 74.28% in Group3. The following Fig. 21 can show these results.

[3] http://socialcomputing.asu.edu/datasets/BlogCatalog3
[4] https://www.kaggle.com/mchirico/cheltenham-s-facebook-group

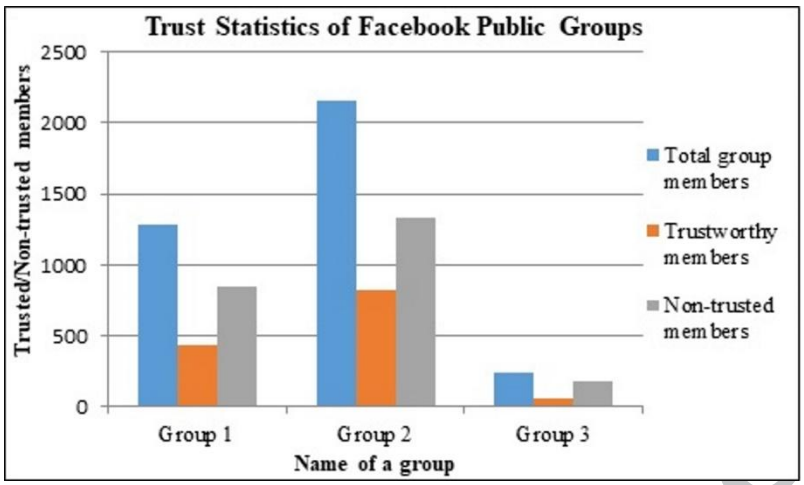


Fig. 21: Results of Indirect Trust calculation

Other findings show that there are 4.5% inactive users of the total members in Group2 who were not sharing posts more than an average post per user in a group and also getting no responses to their posts. Moreover, 5.5% users in Group1 and 6.1% of Group3 were getting no response to their posts. The users of Group2 were more participative on a group level as compared to Group 1 and Group 3. Further, in Group3 maximum 74.28% users were found non-participative. Table 10 represents the results.

Table 10: Summary of Facebook dataset results

Metrics:	**Group 1**	**Group 2**	**Group 3**
No. of in-active member in a group (using Equation 11 and Equation 16)	71	99	15
No. of trustworthy members getting max response, but $participation\ trust < threshold$	7	13	0
No. of trustworthy members getting no response, but $participation \geq threshold$	5	5	0
No. of members having $participation\ trust < threshold$ in a group (using Equation 11)	857	1341	182
No. of members having $participation \geq threshold$ in a group	430	814	63
Avg. Participation trust of a user in a group	0.33	0.38	0.26
Avg. Response trust of a user in a group	0.43	0.45	0.40
Avg. Group level trust of a user in a group	0.38	0.41	0.33

5.3 Discussion

The analysis of trust and influence of various datasets helps us to successfully provide the answers in the context of the stated research objectives. The proposed *SNTrust* model can identify the trusted and non-trusted nodes in the network. The use of Closeness centrality, Eigenvector centrality, and Page Rank algorithm help in finding the influential nodes in the network. The findings from the investigation on all three datasets (mentioned in section 4.1) indicate that the relationship between trust and influence is linearly positive and statistically significant. The results of Pearson's correlation and *P*-value for each dataset remain consistent. The strength of the relationship between trust and influence of a node is found moderate but statistically significant. The association between trusted nodes on a network and community level is found strong. Furthermore, the strength of the relationship between the influence of a node on a network and community level is also strong. Except for the answers to our research questions, it is observed that the trusted nodes are more likely to reside in highest k-core of the network. The direct trust between members of a community is found proportional to the number of triads, number of dyads, the closeness of the members, and average eigenvector centrality found in that community. Likewise, it is observed that

the attributes and interests similarity of users contribute to the trust and closeness between them. The results from the indirect calculation highlight the trusted members in three Facebook communities. Participation of users in terms of number of posts on a group level and actions on those posts from other group members can be examined to find the behavioral patterns of users in Facebook groups.

Generally, the efficiency of an algorithm can be determined based on its usage of different resources (i.e. the cost of time and space used for computation). The SNTrust model computes trust for each node in the network, which is considered as a source of maintaining connections with others nodes (De Nooy et al., 2005). The proposed model is based on the connections of a user in terms of their relationships, communication frequency, degree prestige to calculate its direct trust and on the number of posts of a user in a group and the responses of other group members on that post. Its efficiency is more likely to depend upon the number of connections of a user with others. If a user has less number of connections with other users, then less time will be consumed to calculate his trust value and vice versa. In the same way, higher the number of connections between network users would result in higher computational cost in terms of time. In this perspective, we consider two basic measures of social network analysis i.e. average degree (average connections) of all network vertices and network density for an instance, to clear our point of view to highlight the association of a node's connections with the efficiency of the model. As an illustration, the following Fig. 22 represents that computational time of direct trust for all nodes in the network for Consulting Company Dataset and Freeman EIES dataset. As discussed earlier, the former dataset has more number of nodes (46 nodes) as compared to the latter (32 nodes). However, it can be seen that the execution time for the trust calculation of each node for former dataset is less than the execution time for latter dataset.

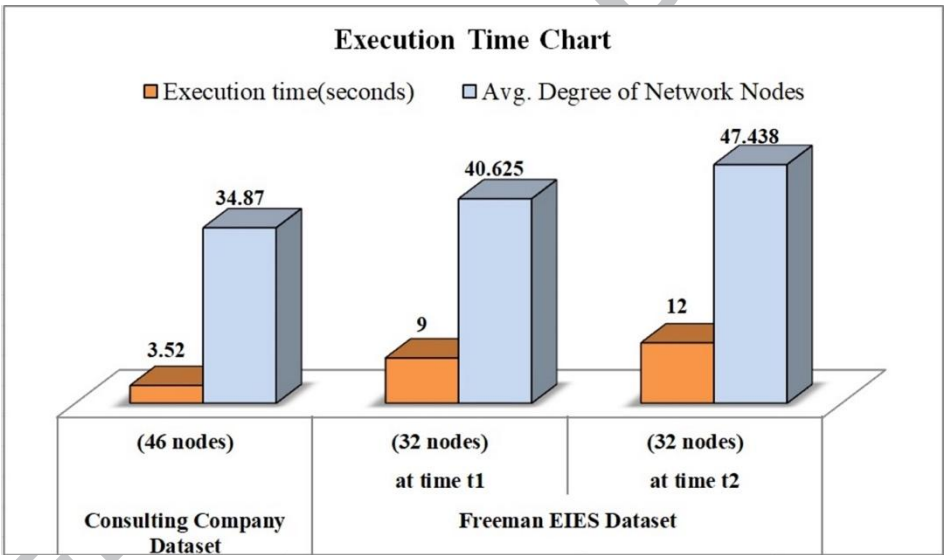


Fig. 22: Execution time for Direct Trust Calculation

It shows that higher number of nodes in the network does not mean that efficiency of model will be less. It implies that efficiency is not likely to be dependant on the number of nodes in this case. Also, it is clear that the average degree of the nodes in the network is positively related to execution time of the model. In other words, if the average number of connections of all network nodes increases, then the execution time also increases and vice versa. Likewise, the value of network density is found maximum for Freeman EIES dataset at time t2 i.e. 0.765 which shows that 76.5% of all possible arcs are present in the network. It shows that in the case of higher number of arcs between nodes in the network, more time is needed for trust calculation for that network. Furthermore, at time t1, in the Freeman EIES network there are 650 ties between 32 researchers and after communication between researchers, at time t2, the number of ties have increased i.e. 749 ties. It is important to discuss that execution time has increased from 9 seconds to 12 seconds with the increase of ties at time t2. Likewise, in the case of indirect trust, the number of posts per user and number of responses against it, would contribute towards more computational costs. Therefore, it is inadequate to specify the exact number of nodes for which the given model can work properly. Because, if there is a small but complete network where each node is connected to every other node in the network then trust calculation will take time. On the other hand, a large but sparse network can take comparatively less time for trust computation.

Besides, efficiency is likely to depend upon the type of trust being calculated. For example, prestige trust (just by counting negative of positive ties from the neighbors of a node towards that node) needs less time to calculate as compared to relationship trust calculation (by giving weights to each relationship of a user with its neighbors and then taking an average value). If we want to calculate trust of user after regular time interval, then the efficiency of the model will depend upon the rate of change in different trust type values of a user. For example, the most changing trust type with respect to the real-world is conversational trust due to the frequency of communication between a user and his connections. On the next level, the values of relationship trust and attribute trust will change respectively, where the attributes of a user and his connections rarely change as compared to relationships between users. So, in this case the efficiency of SNTrust would usually depend upon the type of trust being calculated. Moreover, there are a number of factors which can be considered to enhance the efficiency of an algorithm. First of all, if there are sub-tasks in an algorithm then parallel processing machine can do computations simultaneously (each sub-task can be assigned to a different processor) which can reduce the overall execution time. In the case of SNTrust model, if different types of trust are calculated in parallel by different processors, then it can speed up the trust calculation process. Secondly, the availability of faster, and cheaper hardware with a huge capacity of RAM and hard disk has reduced the problem of handling of large networks in terms of number of nodes and density. However, in this research, the experimental evaluation of SNTrust model is performed on a machine with core i7-7500 processor @ 2.7GHz, 1TB hard disk, and 8GB RAM. However, due to exponential growth of SN users, scalability issue is quite challenging. In this case, computational complexity will exceed for running our trust model to calculate the direct and indirect trust of a user. It is noteworthy that handling the size of data and computational costs in this context is our forthcoming work.

6 Conclusion

This paper provides relative analysis of a node's importance by using two different metrics namely, trust and influence. It highlights the relationship between variables at both the network and the community level. A novel trust model *SNTrust* is proposed to identify trusted and non-trusted nodes in the network. It can help in finding local trust, group-level trust and network-level trust of a user. With the help of results, it can be said that mostly trusted nodes in the network are also influential. The nodes that are trusted/influential at the network level are also trusted/influential in their own communities. Higher the trust among community members, the higher is the cohesiveness between them, which keep them closer to each other. The nodes of a community are more likely to trust each other if they belong to same geographical locations. It entirely depends upon the preference/need of a user whether he/she wants to focus trusted nodes or influential nodes. This study provides a motivation for researchers to narrow down the findings and to empirically find the direction of the relationship (causation) between trust and influence. In the future, we want to work on an adaptive trust model and to use the findings from this research in classification tasks. We shall also try to optimize the efficiency of SNTrust model for multi-processor scenario and to use it in big data concept.

References:

Adali S, Escriva R, Goldberg MK, Hayvanovych M, Magdon-Ismail M, Szymanski BK, Wallace WA, Williams G. Measuring behavioral trust in social networks. Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on. Vancouver, BC, Canada, IEEE; 2010. pp. 150-152.

Araujo T, Neijens P, Vliegenthart R. Getting the word out on Twitter: the role of influentials, information brokers and strong ties in building word-of-mouth for brands International Journal of Advertising 2017;36:496-513.

Asim Y, Niazi MA, Raza B, Malik AK. Personal vs. know-how contacts: which matter more in wiki elections? Complex Adaptive Systems Modeling 2018;6:1-19.

Bae J, Kim S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness Physica A: Statistical Mechanics and its Applications 2014;395:549-559.

Baek S, Kim S. Trust-based access control model from sociological approach in dynamic online social network environment The Scientific World Journal 2014;2014.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks Journal of statistical mechanics: theory and experiment 2008;oct 2008:P10008.

Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations Social networks 2001;23:191-201.

Burt RS. The social structure of competition Explorations in economic sociology 1993;65:103.
Caverlee J, Liu L, Webb S. The SocialTrust framework for trusted social information management: Architecture and algorithms Information Sciences 2010;180:95-112.
Chen D-B, Gao H, Lü L, Zhou T. Identifying influential nodes in large-scale directed networks: the role of clustering PloS one 2013;8:e77455.
Cho J-H, Chan K, Adali S. A survey on trust modeling ACM Computing Surveys (CSUR) 2015;48:28.
De Nooy W, Mrvar A, Batagelj V. Cohesion. In: Granovetter M editor, Exploratory Social Network Analysis with Pajek. Cambridge, Uk, Cambridge University Press; 2005. pp. 59-82.
Gao S, Ma J, Chen Z, Wang G, Xing C. Ranking the spreading ability of nodes in complex networks based on local structure Physica A: Statistical Mechanics and its Applications 2014;403:130-147.
Golbeck J. Trust and nuanced profile similarity in online social networks ACM Transactions on the Web (TWEB) 2009;3:12.
Hamdi S, Gancarski AL, Bouzeghoub A, Yahia SB. IRIS: A novel method of direct trust computation for generating trusted social networks. Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on. Liverpool, UK, IEEE; 2012. pp. 616-623.
Hamdi S, Gancarski AL, Bouzeghoub A, Yahia SB. TISoN: Trust Inference in Trust-Oriented Social Networks ACM Transactions on Information Systems (TOIS) 2016;34:1-17.
Han M, Yan M, Cai Z, Li Y, Cai X, Yu J. Influence maximization by probing partial communities in dynamic online social networks Transactions on Emerging Telecommunications Technologies 2017;28.
Hargittai E, Fullerton L, Menchen-Trevino E, Thomas KY. Trust online: Young adults' evaluation of web content International journal of communication 2010;4:27.
Heidemann J, Klier M, Probst F. Identifying key users in online social networks: A pagerank based approach. Proceedings of the International Conference on Information System. St. Louis, Missouri, USA; 2010.
Jøsang A, Ismail R, Boyd C. A survey of trust and reputation systems for online service provision Decision support systems 2007;43:618-644.
Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, U.S.A, ACM; 2003. pp. 137-146.
Khadangi E, Bagheri A. Presenting novel application-based centrality measures for finding important users based on their activities and social behavior Computers in Human Behavior 2017;73:64-79.
Kiss C, Bichler M. Identification of influencers—measuring influence in customer networks Decision Support Systems 2008;46:233-253.
Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. Identification of influential spreaders in complex networks NaturalPhysics 2010;6:888-893.
Kuan H-H, Bock G-W. The collective reality of trust: An investigation of social relations and networks on trust in multi-channel retailers. ECIS 2005 Proceedings. Regensburg, Germany; 2005. p. 13.
Lewis JD. Trust as a social reality Social Forces 1985;63:967-985.
Li Q, Zhou T, Lü L, Chen D. Identifying influential spreaders by weighted LeaderRank Physica A: Statistical Mechanics and its Applications 2014;404:47-55.
Liu S, Jiang C, Lin Z, Ding Y, Duan R, Xu Z. Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach Information Sciences 2015;306:34-52.
Lü L, Zhang Y-C, Yeung CH, Zhou T. Leaders in social networks, the delicious case PloS one 2011;6:e21202.
Lü L, Zhou T, Zhang Q-M, Stanley HE. The H-index of a network node and its relation to degree and coreness Nature communications 2016;7:10168.
Probst F, Grosswiele L, Pfleger R. Who will lead and who will follow: Identifying Influential Users in Online Social Networks Business & Information Systems Engineering 2013;5:179-193.

Resnick P, Kuwabara K, Zeckhauser R, Friedman E. Reputation systems Communications of the ACM 2000;43:45-48.
Resnick P, Zeckhauser R. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. The Economics of the Internet and E-commerce, Emerald Group Publishing Limited; 2002. pp. 127-157.
S. Nepal WS, and C. Paris. . Strust: A trust model for social networks., International Conference on Trust, Security and Privacy in Computing and Communications. Changsha, China, IEEE Computer Society; 2011. pp. 841-846.
Sabidussi G. The centrality index of a graph Psychometrika 1966;31:581-603.
Sheikhahmadi A, Nematbakhsh MA, Shokrollahi A. Improving detection of influential nodes in complex networks Physica A: Statistical Mechanics and its Applications 2015;436:833-845.
Sheikhahmadi A, Nematbakhsh MA, Zareie A. Identification of influential users by neighbors in online social networks Physica A: Statistical Mechanics and its Applications 2017;486:517-534.
Sherchan W, Nepal S, Paris C. A survey of trust in social networks ACM Computing Surveys (CSUR) 2013;45:47.
Sherchan W, Nipal, S., Paris, C. A survey of trust in social networks. ACM Computing Surveys 2013;45.
Taylor RK. Marketing strategies: Gaining a competitive advantage through the use of emotion Competitiveness Review: An International Business Journal 2000;10:146-152.
Trusov M, Bodapati AV, Bucklin RE. Determining influential users in internet social networks Journal of Marketing Research 2010;47:643-658.
Volakis N. Trust in Online Social Networks. . MSc., University of Edenburg; 2011.
Wang S, Liu Q. Trust-based Access Control in Virtual Learning Community Integration and Innovation Orient to E-Society Volume 2 2007;252:514-520.
Wang Y, Cong G, Song G, Xie K. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. Washington, DC, USA, ACM; 2010. pp. 1039-1048.
Wang Z, Du C, Fan J, Xing Y. Ranking influential nodes in social networks based on node position and neighborhood Neurocomputing 2017;260:466-477.
Xu K, Li J, Song Y. Identifying valuable customers on social networking sites for profit maximization Expert Systems with Applications 2012;39:13009-13018.
Xu W, Lu Z, Wu W, Chen Z. A novel approach to online social influence maximization Social Network Analysis and Mining 2014;4:1-13.
Yang Y, Xie G. Efficient identification of node importance in social networks Information processing & management 2016;52:911-922.
Yap HY, Lim T-M. Social Trust: Impacts on Social Influential Diffusion International Journal of Web Information Systems 2017;13.
Zeng Y, Chen X, Cong G, Qin S, Tang J, Xiang Y. Maximizing influence under influence loss constraint in social networks Expert Systems with Applications 2016;55:255-267.
Zhang Y, Wang Z, Xia C. Identifying key users for targeted marketing by mining online social network. Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on, IEEE; 2010. pp. 644-649.
Zhao X, Liu Fa, Wang J, Li T. Evaluating Influential Nodes in Social Networks by Local Centrality with a Coefficient ISPRS International Journal of Geo-Information 2017;6:35.
Zhao Y, Li S, Jin F. Identification of influential nodes in social networks with community structure based on label propagation Neurocomputing 2016;210:34-44.

Highlights:

- The analysis of trust and influence of a node is performed on the network level and community level.
- *SNTrust* model is proposed to find trusted/non-trusted nodes in a network.
- The standard methods for influential node's detection are applied on a number of standard datasets to find influential nodes in the network.
- The relationship between trust and influence is investigated by using Pearson's correlation and P-value.