Expert Systems With Applications 122 (2019) 1–15

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Discover opinion leader in online social network using firefly algorithm

Lokesh Jain, Rahul Katarya*

Department of Computer Science & Engineering, Delhi Technological University (Formerly Delhi College of Engineering), Shabad Daulatpur, Main, Main Bawana Road, Delhi 110042, India

A R T I C L E I N F O

Article history:
Received 11 June 2018
Revised 21 December 2018
Accepted 22 December 2018
Available online 23 December 2018

Keywords:
Online social network
Opinion leader
Firefly algorithm
Community partitioning

A B S T R A C T

Nowadays, with the widespread access to web 2.0, the social network plays an unbelievable role in knowledge sharing and diffusion of new products. People can share their views and can visit other's opinion about the particular material, news, products, artifacts and, trends, etc. anywhere, anytime, and anywhere. An Opinion leader is a critical person who can change, modify and transform other's view by their knowledge and proficiency. In this article, an innovative approach is proposed to discover the top-N local and global opinion leader within the community and social network respectively. Initially, we identified the community structure within the social network using the modified Louvain method and next identified the opinion leader using a modified firefly algorithm in each community. We also determined the global opinion leader within the same social network using the same firefly algorithm. The proposed approach is exceptionally supportive to expert and intelligent system because it competently discovered the local optimum concurrently in each subgroup of the social network. All the users can update its attractiveness value without any supposition, and as soon as the distance among the user's increases, the other users can automatically create another subgroup in the network and form the local community. In addition, as the population size in the network increases, the entire users measure their prominence simultaneously. Therefore, there is no consequence on computational time and accuracy of the algorithm. Thus, the proposed algorithm is superlative suitable for discovering the opinion leader in the local community and globally in the social network. For legalized the proposed approach, we implemented our proposed method on synthesized as well as on real dataset. Finally, we concluded that both the recommended procedures are much better concerning the accuracy, precision, recall, and F1-score with the widely used standard Social Network Analysis (SNA) measures.

1. Introduction

In the current web scenario, social networking is an integral part of human life. These sites, such as Facebook, Twitter, Tumblr, and Instagram, etc. provide the opportunity to interact with another unknown world of known things and human being. Currently, these Websites are the primary source of information transmission and dissimilation (Boyd & Ellison, 2007; Scott, 2000; Wasserman & Faust, 1994). Moreover, social networking sites often provide a platform for the companies for the diffusion of the new product and merchandise (Abrahamson & Rosenkopf, 1997; Bakshy, Rosenn, Marlow, & Adamic, 2012). The success rate of this diffusion is dependent upon identifying the critical users effectively in the social network, such type of person known as an opinion leader. Opinion leader has more incredible significance for the diffusion of the new product. Opinion leader also affects the consumer behavior and decision making by their knowledge and experience about a particular product (Chan & Misra, 1990; Myers & Robertson, 1972).

According to Dye (2000), an opinion leader is a person or set of persons having more influence on the customer's adoption process and decision making. Lazarsfeld et al. have introduced the phenomenon of the opinion leader in their seminar in 1940 and 1950 (Gold, Katz, Lazarsfeld, & Roper, 1956). According to them, identification of opinion leader is a two-step procedure in which in the first step, opinion leader analyzes, examine and understand the end user's requirements, and in the second step, opinion leader derives their own opinion from the first step incorporated with their knowledge and skills.

We introduced the problem of identification of local and global opinion leader in the online social network. Many researchers have

* Corresponding author.
E-mail addresses: nsitlokesh.jain@gmail.com (L. Jain), rahulkatarya@dtu.ac.in (R. Katarya).
URL: http://www.dtu.ac.in/Web/Departments/CSE/faculty/ (R. Katarya)

https://doi.org/10.1016/j.eswa.2018.12.043
0957-4174/

made various efforts and studied to find the solution to this problem. To solve this problem, initially, we partitioned the social network in multiple communities and next we, attempted to identify the opinion leader. We have also used the firefly algorithm to determine the global and local opinion leader. The Firefly Algorithm (FA) is a naturally inspired meta-heuristic optimization algorithm (Liu, Tian, Zhang, Yuan, & Xue, 2013). The concept of the algorithm is based on the global community and blinking actions of Firefly. In this technique, the flashing light of fireflies helps to find out the mates, enticing the potential target and protecting themselves from their enemies and hunters. The same approach we have applied for the identification of opinion leader. The leading cause behind implementing FA is that it is best suited for our model due to its searching optimality. For finding the local optimal in local communities and global optimal in the social network, it works in both exploitation and exploration phase respectively. Therefore, both the searching strategies are too significant in their respective region. Another advantage is that the value of the control parameters can be changed to minimize the total number of iterations if it converges. Besides, the FA is also proficient to handle a large number of nodes because as the nodes increases, it locates the local optimal simultaneously in each community; therefore, the overall complexity does not change as the size of the network changes. The inherent characteristics of the individual entity include centrality, rank, and prestige that attract other objects in the social networks. These characteristics not only entice the other actor but also able to change the behavior and decision of another user.

The proposed model behaves reminiscent of an intelligent expert system to classify the user as opinion leader by using its optimal decision-making capability in the social network. An Expert system is a system that imitates the human decision making knowledge and behavior in a particular domain. For classification purpose, it utilizes the nature-inspired firefly algorithm in which all the heuristic control parameters update automatically until the result converges. The knowledge base of the model is the degree of trust, and various centrality measures about the users in the network and the collected knowledge are used to measure the attractiveness of the user. Therefore, the proposed model is capable of identifying the highly knowledgeable, trusted, connected, socially affluent, and expert people without any human inferences.

The pseudo structure of the whole paper is as follows: in the first segment, we dealt with the literature work related to the opinion leader identification. Various methods have proposed for identification of opinion leader by the different studies in the different domain of interest. In the second segment, we present the problem origination and the background details of the social network. In the third segment, we gave our proposed methodology based on firefly algorithm. We also presented how the heuristic parameters are elected, the complexity of the proposed algorithm and motive for preferring firefly algorithm in detail. In the fourth segment, we demonstrated the experimental result with a synthesized and read data set. In the fifth segment, we described the strengths and weaknesses of the proposed method and in the last section; we have discussed the summary, limitations and future scope of the proposed plan. Therefore, the primary vital features of the article are as fol1lows:

- A novel and innovative approach suggested that classify the opinion leader not only at the global level but in the local community also.
- Nature inspired meta-heuristic approach, using the firefly algorithm, is anticipated to discover the opinion leader.
- An improved and modified Louvain method proposed to find out the communities within the social network.
- The overall complexity of the projected firefly algorithm is superior to the previously developed standard Social network Analysis (SNA) measures in term of accuracy, precision, recall, F1-score, and computational time.

2. Related work

The influence of the user has defined in many ways in the social network. Various studies used the different approaches to find out the opinion leaders. Some of the methods included the degree of trust, trust propagation, game theory, node centrality, user status, user activities, response time, text mining, sentiment analysis and many more as a significant aspect to identify the same. According to Trusov, Bodapati, and Bucklin (2010), the user's total number of activities and levels influenced the behavior and decisions of their friends. They proposed a variation of Bayesian shrinkage in Poisson regression to determine the significant effect of user activity using the user's activity records.

Li and Du (2011) proposed an ontology-based model BARR, which included blog content, authors, readers, and their relationship, to identify the hot topics. Further, the identified hot topics linked with opinion leader, and the decision makers used the content posted by opinion leader on these hot topics, for significant marketing policies. According to Mak (2008) a game theoretic based approach is used that determines the null or weak result on the association of opinion leader-follower. The projected model consists of customer heterogeneity, Word-of-Mouth messages, time fondness, and social network. Goyal, Bonchi, and Lakshmanan (2008) projected a different method to discover the opinion leader with frequent pattern mining approach where a table of user actions created and analyzed their actions. Aghdam and Jafari Navimipour (2016) designed a new framework based on user's total trust value (TTL) and the opinion leaders identified based on highest TTL. Cho, wang, and Lee (2012) proposed a method to recognized opinion leader as a top marketing alternative in term of dispersion speed and a supreme cumulative number of adopters. They also predicted that the opinion leaders are better for the wild dispersion and have the highest sociality.

Ma and Liu (2014) proposed super network theory, which was a summation of text mining and network topology analysis and defined four parameters: node superdegree, superedge degree, superedge–superedge distance, and superedge overlap for identification of opinion leader. Van der Merwe and Van Heerden (2009) proposed a hypothesis and challenged the postulate that opinion leaders are topic specific by representing the strongly coupled relationship between the domain-specific opinion leadership with general opinion leadership. Li, Ma, Zhang, Huang, and Kinshuk (2013) proposed a mixed framework that evaluated text contents and time influenced user behavior. The top-N opinion leaders identified based on four attributes: novelty, activity, expertise, and influence. After this step, the performance of the opinion leader evaluated concerning the parameters longevity and centrality. Bodendorf and Kaiser (2010) anticipated a model built on the concept of text mining and social network analysis. In this model, all the users who worked on the same type of text in the social network are identified. Further, opinion leaders were discovered based on higher social relations. Shafiq, Ilyas, Liu, and Radha (2013) proposed an innovative Longitudinal User Centered Influence (LUCI) model that practiced the user collaboration information, further, with the help of this information, clusters of users made and categorized them into four leader's classes. They validate their experiment by implementing it on Everything2 social network dataset. Song, Chi, Hino, and Tseng (2007) proposed a new algorithm, called InfluenceRank, that ranked blog not only considering the content of the blog but also considered the involvement of the blog in the network. They also stated that opinion leaders propagated new ideas, innovations, and opinions in the blogosphere network. Aleahmad, Karisani, Rahgozar, and Oroumchian (2016) proposed an algorithm, called OLFinder, classified the core topic of argument in social network and computed the capability and the status notch of each actor in that domain. After calculating these measures, a rank list of opinion leaders obtained in the particular domain. Heidemann, Klier, and Probst (2010) proposed a PageRank based approach that considered the user's connectivity and communication activity as parameters. Sharara, Getoor, and Norton (2011) projected a probabilistic surveying method that united the secondary source of data with the partial primary data. Opinion leaders discovered based on the network of influence, i.e., consider only those users who were influenced by the opinion leaders. Zhu, Lin, Lu, and Wang (2016) anticipated a method based on sentiment analysis by examining user's emotional preferences and social network structure. Duan, Zeng, and Luo (2014) also proposed a framework based on sentiment analysis in a web-based stock message board. Initially, they computed the user activity features based on the posts and then clustering techniques applied to generate clusters with probable opinion leaders. Chen, Hui, Wu, Liu, and Chen (2017) designed a novel system, called D_OLMiner that identified the opinion leader in a dynamic social network by considering the time constraints. Luo, Yang, Chen, and Wei (2018) also proposed an improved weighted LeaderRank algorithm based on a far-reaching preliminary effect of users and the total number of user interactions.

Consequently, the variety of researchers proposed diverse models, algorithms and framework in the distinct domain of interest to understand the behavior and decision of people in the social network from mathematics (Bonacich, 1972; Freeman, 1978b), e-commerce (Zhao, Kou, Peng, & Chen, 2018), business and marketing (Dichter, 1966; Domingos & Richardson, 2001), and computer science (Page, Brin, Motwani, & Winograd, 1998; Sharara et al., 2011). The comparative studies of most of the previously proposed approaches are as follows:

Author, year	Method	Mechanism	Pros	Cons
Song et al., 2007	Influence Rank algorithm	Identify the leader based on the content posted by the user in the blogosphere. The algorithm also ranked the blogs according to the contribution in the network.	Simple and uses coverage, diversity, and distortion metrics.	Only applicable in the blogosphere; does not describe the method to remove extraneous content.
Mak, 2008	Model-based on game theoretic approach	Determined the null or weak result on the association of opinion leader-follower as well as also combined the world of mouth and customer heterogeneity with the social network.	Included the time discount factor and provided equilibrium in the network.	Highly sophisticated and composite.
Goyal et al., 2008	Framework using frequent pattern mining approach	Maintained the table of user's events and analyzed their actions. Identified the opinion leader based on the trends that are highly followed by their followers.	Scalable and highly efficient.	Highly complex; does not support by all the datasets.
Van der Merwe, R.; van Heerden, 2009	Domain-specific opinion leadership hypothesis	Link the leadership phenomenon with the social network theory and proposed that opinion leader are more domain specific rather than topic specific.	Useful for marketing strategies and the diffusion of new products.	Limited only for marketing area and few centrality measures used.
Trusov et al., 2010	Bayesian shrinkage in the Poisson regression model	Determined the significant effect of user activity using the user's actions records.	Identify only those users who influence user's activity.	Non-standard complex system.
Heidemann et al., 2010	PageRank based approach	Analyzed the user's connectivity and communication activity as parameters and mixed it with design science research.	Better precision and accuracy; useful for advertising strategies.	Specific for marketing; parameters are not clearly defined.
Bodendorf & Kaiser, 2010	Design a model that supports text mining in social network	Discovered the opinion leader in the online forum using text mining and determined the opinion trends based on the relationship with other users who also works on same text.	Simple; use of tweet text.	Suitable only for online forum and community.
Sharara et al., 2011	An algorithm using the probabilistic approach	The method combined the secondary source of data with the partial primary data. Opinion leaders discovered based on a network of influence.	Intelligently used primary and secondary data; reduce association cost.	Missing a full joint approach, do not use the weighted mechanism.
Li & Du, 2011	An Ontology-based model called BARR	Include blog material, novelist, lovers, and their relationship to find the lighted area. Opinion leaders selected based on the material placed by them in the blog.	Useful for marketing strategies.	Only consider the blogs.
Cho et al., 2012	Framework Used dispersion speed and the supreme cumulative number of adopters	Opinion leaders discovered from high sociality that is measured using dispersion speed and the total number of adopters.	Useful for marketing strategies, diffusion of new products, and cascading behavior.	Does not work if the initial amount of adopters is missing or unknown.

Li, Ma, Zhang, Huang, and Kinshuk, 2013	Mixed framework	Initially, evaluated text contents, time and user behavior in online learning community then opinion leader discovered based on four attributes: novelty, activity, expertise, and influence.	High performance and effortless.	Only topic specific opinion leaders identified.
Shafiq et al., 2013	Longitudinal User Centered Influence (LUCI) Model	Maintain user interaction information and classify the users into four leader's categories. Validate the result on two real data sets.	Use topological attributes of the network; high accuracy for a limited data set.	Does Not support scalability; does not suitable for the dynamic social network.
Ma & Liu, 2014	Super network based Theoretical algorithm	The proposed algorithm is a summation of text mining and network topology analysis. The authors also introduced four new super network indexes to rank super edges in the multidimensional model.	Uses the text mining and network topology and Include all common, psychosomatic, and ecological factors.	Complex and not suitable for all type of social network.
Duan et al., 2014	Framework using sentiment analysis	Mixing the clustering algorithm with user sentiment and computed the user activity features based on the posts. Clustering techniques applied to generate clusters that enclosed effective opinion leaders.	Useful for web-based store communication board; less complex	Only applicable in an online message board; does not describe the method to remove garbage content from tweets.
Aghdam & Jafari Navimipour, 2016	Framework using user's total trust value (TTV)	Computed the TTV for all users in the network and chosen the set of users having the highest TTV.	Simple, Suitable social network marketing (SNM) campaigning.	Does not support by all dataset; lesser accurate.
Aleahmad et al., 2016	OLFinder algorithm	Categorized the core topic of argument and computed the capability and the status value of each actor in that domain.	Accuracy in average precision and recall; simple algorithm.	Use the number of re-tweets to measure the user's reputation.
Zhu et al., 2016	Leader-PageRank model based on Sentiment analysis	Examined user emotional preferences and social network structure. An edge weight matrix designed to discover opinion leader	More accurate and consider human emotions.	Complex; does not include all emotions.
Chen et al., 2017	D_OLMiner algorithm	Constructed the dynamic social network, found the communities and measured the centrality of each user.	Reduced computation time; solve the overlapping influence problem.	Complex; does not suit for all dynamic social network

Most of the discussed approaches applied on static social networks in which the researchers considered the social network as the invariant structure and find the opinion leader at the global level. We have proposed a variety of algorithm in which opinion leader identified at the local level within the community and the global level in the network using the meta-heuristic firefly algorithm.

3. Problem origination

In this segment, we discuss some familiar concepts related to the graph theory to formulate our problem. Social network epitomizes the social structure organized around the group of actors. The social network can be well-defined as per the un-directed graph $G=(V, E)$ where V signifies the group of nodes indicating actors or user and E means the group of edges indicating the relationship between the actors. The relationship between the actors can be friendship, acquaintances, author- coauthor relationship and many more. Centrality component of a node is suitable to classify the most convenient and influence node in the network (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Opsahl, Agneessens, & Skvoretz, 2010; Ruhnau, 2000). Furthermore, in the section, we discuss the four types of different centrality measures: closeness centrality (CC), betweenness centrality (BC), eigenvector centrality (EC), and PageRank (PR) (Freeman, 1978a), used for calculating the prominence P of the node that we will confer in the next part of the section.

3.1. Closeness centrality (CC)

This degree grooves a value for each node based on their closeness to all other nodes within the network. This scheme computed the shortest paths between all nodes and based on the sum of the shortest path; assign a value to every node as shown in Eq. (1).

$$\mathrm{CC(x)} = \frac{1}{\sum_{\mathrm{y=1}}^{\mathrm{y=n}} \mathrm{d}_{\mathrm{(x,y)}}} \tag{1}$$

Where $d_{(x,y)}$ is the shortest distance between node x and node y.

3.2. Betweenness centrality (BC)

Betweenness centrality considers the degree that counts the occurrence of a node on the straight path between other nodes. It can be defined as the fraction between the total number of shortest paths exists between node i and node j passes through node x to the total number of shortest paths exists between node i and node j as shown in Eq. (2).

$$\mathrm{BC(x)} = \sum_{\mathrm{i\neq x\neq j}} \frac{\mathrm{c_{(i,j)}\ (x)}}{\mathrm{c_{(i,j)}}} \tag{2}$$

Where $c_{(i,j)}(x)$ represents the total number of the shortest paths between node i and node j passes through node x, and $c_{(i,j)}(x)$ represents the total number of the shortest paths between node i and node j,

3.3. Eigenvector centrality (EC)

Eigenvector centrality is used to quantify the impact of a node in the network. In this scheme, all nodes having a relative mark constructed on the knowledge that association to high-marking nodes contribute more to the score of the node than parallel connections to low-marking nodes. This measure shows the idea that an actor is all the more vital as it associated with actors who are themselves essential as shown in Eq. (3).

$$\mathrm{EC(i)} = \frac{1}{\lambda} \sum\nolimits_{j \neq i} \left(y_{ij}.x_j \right) \tag{3}$$

Where y_{ij} represents the adjacency matrix and node j is the neighbor of node i.

3.4. PageRank (PR)

PageRank algorithm used to uncover influential or important nodes whose influence extends beyond just their straight acquaintances as shown in Eq. (4). This algorithm used in citation network, monitoring network activity, etc.

$$\mathrm{PR(x)} = \alpha + \sum\nolimits_j a_{ij} \frac{xj}{L(j)} + \frac{1-\alpha}{N} \tag{4}$$

Where *L(j)* is the total number of neighbors of node j.

Now the factor additive centrality θ_x of node x can be computed as shown in Eq. (5).

$$\theta_x = \frac{\mathrm{mean}\ (CC(x), BC(x), EC(x), PR(x))}{d_x} \tag{5}$$

Where d_x represents the degree centrality of node x. The value of additive centrality θ_x is used to compute the prominence of a node, and it diverges between 0 and 1.

3.5. Prominence

According to the proposed approach, Prominence of a node can define as a function that takes various types of centrality as input and produces a discrete value as prominence value for each node. The prominence *P* of a node x is inversely proportional to the additive centrality θ_x of node x as shown in Eq. (6).

$$P_x \alpha \frac{1}{\theta x} \tag{6}$$

If we analyze the network structure enormously, we would observe that the node is having the smaller mean value of additive centrality θ_x but having higher degree centrality. The reason behind this disparity is the stable and unstable triangle in the network, homophily, and triadic closure. Therefore, we characterized the inverse relationship between prominence and additive centrality θ_x by considering the degree centrality as measure feature for computing prominence.

4. Proposed approach

In this segment, we demonstrate our proposed approach to discover the local and global opinion leader in the social network. First, we projected a modified Louvain method to find out the communities in the social network built on the modularity gain of the network. In the next step, we used the modified firefly algorithm to find out the opinion leader in the interclass and intra-class community. In the modified firefly algorithm, firefly indicates the set of users who want to search the person having more attractiveness as compared to their attractiveness. The experimental result suggested that the anticipated algorithm is better than the previously defined standard SNA measures. Now, we discuss the modified community partitioning algorithm and firefly algorithm.

4.1. Community partitioning algorithm

A community can define as sub-network or group of nodes in the network that represent some features of the network. In a social network, it is a very crucial task to identify communities. For example, Communities can be a cluster of users who follow the same religion, group of animals living in the same geographical region or network of author-coauthor who works on the same subject, etc. We have used the classical Louvain method for community detection in the social networks. In Louvain algorithm, a greedy approach uses with hierarchical clustering in which continuously removed the edges with higher betweenness centrality. The central concept of this algorithm is based on the modularity of the network. The value of modularity lies between 1 and −1 which indicate the density of the edges of the network. If the density of the network is very high, all the nodes are coupled with each other strongly otherwise weakly coupled. The Louvain algorithm is separated into two phases alternatively. In the first phase, every node belongs to its community, and in the second phase, every node is measured separately and positioned in the neighboring community having the maximized modularity gain (Clauset, Newman, & Moore, 2004; Newman, 2006). The pseudo code of the modified Louvain community detection algorithm is summarized in Algorithm 1.

The modularity gain can be computed as shown in Eq. (7).

$$M' \frac{1}{2m} \sum\nolimits_{xy} \left[A_{xy} - \frac{k_x k_y}{2m} \right] \delta(c_x c_y) - (1 - C_x) \tag{7}$$

In the above equation, c_x and c_y is the community of node x and node y respectively. The worth of function δ will be one if both nodes belong to the same community otherwise it will be zero. The variable A_{xy} represents the weight between node x and node y. The factor $\frac{k_x k_y}{2m}$ represents the expected number of nodes between node x and node y where k_x and k_y is the degree of node x and node y respectively and, m indicates the total weight of the network. The value of modularity gain $\acute{M}$ lies between 1 and −1. In the modified method, we proposed that the clustering coefficient C_x of node x is also affecting the modularity of the community. As soon as the clustering coefficient of a node gradually increases, the probability of the node of being the portion of the same community also increases and vice versa.

4.2. Firefly algorithm's variation in the social network

Xin-She Yang instigated firefly algorithm in late 2007 at Cambridge University. The basic idea of the firefly algorithm based upon the flashing behavior of fireflies. We proposed the variation of the mentioned algorithm for the social network to discover the opinion leaders. In the exception of the suggested algorithm, the user behaves as Firefly and attractiveness of the user is proportional to the prominence of the user. Therefore, our algorithm proposed the following guidelines:

- Initially, the entire user in the social network treated homogeneous user regardless of their gender.
- Additive centrality θ_x of the user is used to compute the prominence *P*.
- The attractiveness β of the user (Firefly) is proportional to the prominence (brightness) of the user and as the centrality (distance) decrease; prominence about a user also decreases due to triadic closure and homophily - the user attempt to discover another user having more prominence than their prominence in a particular domain. If the user could not find the one having high prominence, the user stops the search.
- Ranked the users based on attractiveness score to find opinion leader.

Algorithm 1
Modified Louvain community detection.

Input: Directed social network graph $G=(V, E)$
Output: The hierarchical set of communities G'.
Steps:
1. $G'=\{\}$, the empty set of communities
2. node $n_i \in G$, put node n_i in its local community C;
3. while (j $\neq$ max_no_of_node) do
4. for all node n of G
Position node n in its neighboring community C' has the maximized modularity gain $\acute{M}$.
5. end for;
6. end while;
7. if (new_modularity > old_modularity)
$G'=$the network between communities of G;
8. else
9. stop;
10. end if;

4.2.1. Attractiveness

According to the proposed model, the attractiveness β of the user merely depends upon the light absorption coefficient γ and prominence P of the user. Initially, if the user has not connected to the network, still, the user has gained some attractiveness β_0 based on some prior interactions and experiences with other users. If the user moves far away with the distance d from another user, its attractiveness also decreases exponentially. Hence, we can define the attractiveness β of the user i within the community as shown in Eq. (8).

$$\beta = \beta_0 \mathrm{e}^{-\gamma \mathrm{d}^2} + \mathrm{P_i} \tag{8}$$

Where β_0 is the attractiveness at distance $\mathrm{d}=0$, P_i is the prominence value of user i, and γ is light absorption coefficient.

4.2.2. Progress

The progress x_i^{t+1} of the user i, towards the more attractive user j, can be computed using Eq. (9).

$$\mathrm{x_i^{t+1}} = x_i^t + \beta_0 \mathrm{e}^{-\gamma \mathrm{d_{ij}^2}} \left(x_j^t - x_i^t \right) + \alpha_\mathrm{t} \in_i^t \tag{9}$$

Where x_i^t is the location of user i at time t, the second fraction of the equation exists due to the attraction and the third fraction of the equation, indicates the randomization process with α_t is a randomized parameter, and $\in_i^t$ is a vector of random numbers. When the value of the light absorption coefficient γ is zero, it is equivalent to the Levy flight firefly algorithm.

4.2.3. Search optimality in firefly algorithm

The vital component of any of the nature-inspired algorithm is the searching optimality that indicates how efficiently algorithm obtains the desired result. The firefly algorithm also exhibits the two searching strategies: exploitation and exploration to find the local and global optimal respectively (Fister, Yang, & Brest, 2013). In the proposed approach the exploration is gained by randomization that included the random search by the user in the social network to find the other users having more attractiveness than their self-attractiveness. Therefore this searching approach is beneficial to perceive the optimal opinion leader in the global extent. But for achieving the desired result, the appropriate amount of balance is needed between randomness and universal search. If the randomness is too high, it may be possible that the algorithm converges very soon or neglects some local minima of the subgroups; hence proper tuning required between randomness and universal search.

For implementing the exploitation, many attributes required that consists of information and experience about the local region, the total number of users and their neighbors, and the distance between the users. Besides this information, some other knowledge related to subgroup shapes such as convexities, gradients, and record of past processes also needed to find the local minima. The pragmatic study from the observation states that exploitation tends to enhance the rate of convergence of the algorithm; on the other hand, exploration tends to decrease the rate of convergence of the algorithm.

The harmonizing between the exploitation and exploration is dependent upon the nature of the network and its surroundings. Landscape-based optimality includes the entire information regarding the whole network, total number of users, centrality, in-degree, out-degree, clustering coefficient and many more so that an optimal solution can found at local as well as global level with a lesser amount of exertion concerning time and number of iterations. Although there is no proper guideline for landscape-based optimality, however, this synchronization mainly depends upon the concrete landscape based optimality. In the proposed approach, both exploitation and exploration are used to find the local and global optimum minima.

4.3. Parameters setting

For implementing the social network based firefly algorithm, initially, we set up the heuristic control parameter with the best environment for the proposed research. Initially, we took some random values for the parameters and compared the results accordingly. In our study, not all the settings are heuristic; some static parameters are also existed such as the size of the network, the centrality of the node, and the degree of trust. We considered only 40% of the population for the test set and the remaining 60% of the population used for validating the parameter's value in the research.

In our research work, attractiveness ($\beta 0$), light absorption coefficient γ as prestige, randomize parameter α, the total number of users in the network n, and the maximum number of iterations i are heuristic parameters. The combination of n^*i is appropriate for obtaining the solution space of the problem. If the value of the n^*i factor is immense, there is a privileged probability to achieve the better value for all the parameters; but due to memory, time and resource constraints, we have practically considered only 2500 nodes of the 'Small Slashdot' dataset to hold the calculation of parameter within time. The value of all the parameters is analyzed using the linear model Analysis of Variance (ANOVA) approach (Sthle & Wold, 1989). Initially, we take three groups in which the first group consists of 1000 users that processed over 100 iterations, the second group consists of 2000 users that handled over 50 repetitions, and the third group comprises of 500 users that processed over 2000 iterations. For filling the entire entries in the ANOVA table, we considered the additive centrality of each user for calculation.

The ANOVA model consists of Degree of Freedom (DF), Mean Square (MS), Sum of Squares (SS), F-statistic, and P-test values as

Table 1
Values of heuristic parameters using ANOVA.

n*i	Parameters			Degree of Freedom (DF)	Mean Square (MS)	Sum of Squares (SS)	F-statistic	P-test
	α	β_0	γ					
(500*200)	0.1	0.1	0.01	2	45,893	2,674,981	25.87	0.000
	0.25	0.25	0.1	2	46,348	2,745,896	32.51	0.000
	0.5	0.5	0.15	2	52,872	3,876,403	59.07	0.003
	0.75	0.75	0.2	2	35,789	1,784,975	39.52	0.000
	1	1	0.25	2	27,841	879,423	17.93	0.000
(1000*100)	0.1	0.1	0.01	2	45,893	2,758,106	39.05	0.000
	0.25	0.25	0.1	2	56,489	3,917,393	45.92	0.000
	0.5	0.5	0.15	2	62,958	4,728,491	65.20	0.005
	0.75	0.75	0.2	2	60,013	4,187,928	52.69	0.001
	1	1	0.25	2	41,433	2,711,942	21.72	0.000
(2000*50)	0.1	0.1	0.01	2	16,784	1,078,349	24.97	0.000
	0.25	0.25	0.1	2	27,538	2,187,592	36.80	0.001
	0.5	0.5	0.15	2	38,222	4,007,828	48.53	0.004
	0.75	0.75	0.2	2	32,989	3,335,713	27.41	0.000
	1	1	0.25	2	29,483	2,967,485	15.36	0.000

shown in Table 1. Only those parameters consider as a statistically optimal parameter for which the value of P is less than or equals to 0.05 with 95% confidence intensity.

From the above statistical result, we have concluded that the optimal value for the heuristic parameters attractiveness β_0 is 0.5, value for light absorption coefficient γ is 0.15, and value for randomizing parameter α is 0.5. The heuristic parameter values that are achieved using the ANOVA model are optimal values because these values are suitable for our dataset under few constraints. In our model too, there are few constraints also exists such as each user must know the degree of trust of other users, the initial size of the network must identify, in-degree, out-degree, and clustering coefficient of each node should be known. If the dataset revolutionizes, it may be possible that the obtained parameters values will also diverge. Therefore, these values are the viable values that are used to maximize the attractiveness of the user for our dataset.

4.4. Complexity of firefly algorithm

The complexity of most of the nature-inspired algorithm is uncomplicated and straightforward. The firefly algorithm consists of one outer loop and one inner loop that iterates over the total number of maximum iteration m and the overall size of network n, respectively. So, the worst-case complexity of the algorithm is n^2m (Fister et al., 2013). Although, the complexity of the algorithm is linear if the iteration is very far above the base level and n is relatively near to the base level. The computation of the algorithm also engrosses the calculation of attractiveness value that is a linear function ω, involves the calculation of the degree of trust and centrality of the entire node in the network. The complexity of this fraction is $n\omega$ that is linear; therefore, the overall complexity of the proposed algorithm is $\Theta\ (n^2m + n\omega) \approx O\ (n^2)$. The pseudo code of the modified firefly algorithm is summarized in Algorithm 2.

The firefly algorithm initiated with a set of users called population, and define the light absorption coefficient γ as prestige, for each user in the network. In this case, the user's behavior is not similar to flashing behavior of firefly that releases the light to attract the other fireflies; therefore the initial value of the light absorption coefficient γ computed using the degree of trust that is achieved by the user in the network (Adali et al., 2010; Sherchan, Nepal, & Paris, 2013). The degree of trust T between user x and user y can be computed using the following Eq. (10).

$$T = f(d_{xy}, d_{yx}, d^z_{xy}, d^z_{yx}, r_x, r_y) \tag{10}$$

Where d_{xy} is the degree of trust expends by user x on user y, d_{yx} is the degree of trust expands by user y on user x. d^z_{xy} is the degree of trust that is supposed to be suggested by user x to user y on user z, d^z_{yx} is the degree of trust that is supposed to be indicated by user y to user x on user z. r_x and r_y is the reputation of user x and user y respectively. The trust can be direct trust (DT), indirect trust (IDT), and recommended trust (RT) as shown in Fig. 1.

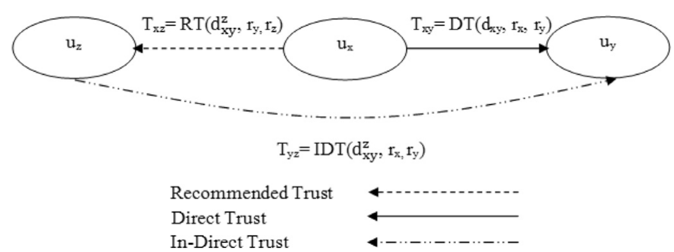


Fig. 1. Trust representation between the users.

Further, each user attempts to search a user having more attractiveness that computed with the help of prominence of the user. In the mentioned algorithm, there are two iterations: the first iteration for the outer users and second iteration for the current user whose prominence value compared with the prominence value of the other users. If a user searches another user having higher attractiveness, the first user updated its knowledge about the attractiveness of user and socialized this knowledge to its all dimension. Similarly, all the users have pursued the identical procedure and socialize the experience to their entire neighbor in all aspects. At last, all the users who have higher attractiveness considered as opinion leaders in the social network. The flow chart of the proposed algorithm represented in Fig. 2.

4.5. Why do we prefer firefly algorithm?

To efficiently answer this question, we deeply analyzed the main feature of the firefly algorithm that makes it so efficient to identify the opinion leader in the social network as follows:

- The firefly algorithm is nature inspired swarm intelligence based heuristic algorithm in which multiple agents interacted with each other and solves the global optimization problems. The central concept of the firefly algorithm based on the brightness and attractiveness of the firefly. As soon as the distance between the user's changes gradually, the attractiveness factor of the user also updated in the same proportion and the whole population of the network is automatically divided into multiple subgroups. In each subgroup, all users move around local optimum. Once the local optimum of all the subgroups has measured, the most excellent global optimum solution can be established (Yang & He, 2013).

Algorithm 2
Modified Firefly Algorithm.

Input:
1. Generate an initial population of users $u_i(i = 1,2, 3...n)$
2. Initial light absorption coefficient γ as prestige.
Output: list of users having the highest attractiveness
Steps:
1. Compute the prestige of each user based on the degree of trust.
2. Set all the heuristic parameters attractiveness β_0, light absorption coefficient γ, randomizing parameter α, derived using the ANOVA model.
3. Evaluate initial attractiveness of the user u_i using Eq. (8).
4. Measure user's progress toward other user using Eq. (9).
5. while (iteration ≤ max_iteration)
{
6. for (i = 0; i < user_population_size; i++)
7. for (j = i; j < user_population_size; j++)
{
8. if (Pi > Pj)
9. Update the user's attractiveness via $e^{-\gamma}$
}
10. Update current user attractiveness
11. iteration++;
}
12. Return the list of users with their attractiveness.
13. end;

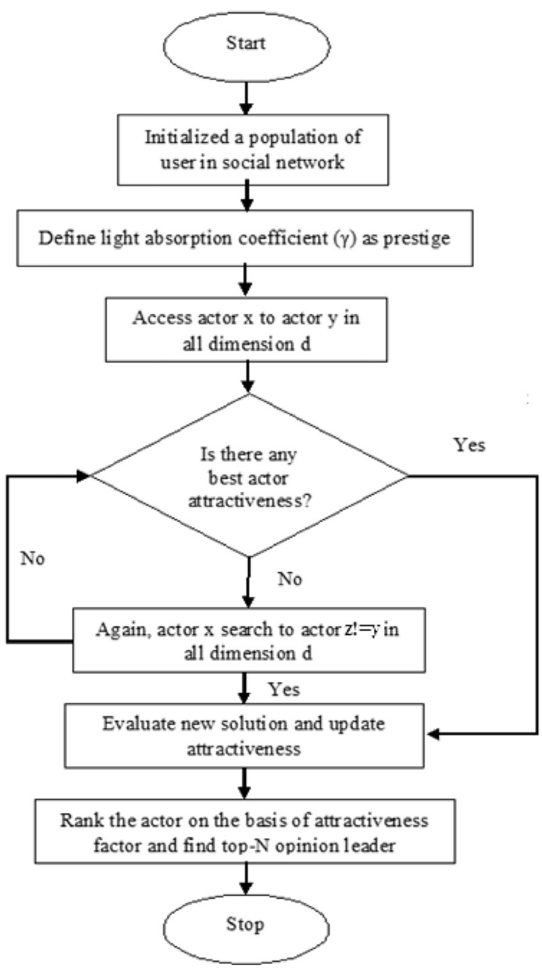


Fig. 2. Firefly algorithm flow chart.

- The second chief virtue of this algorithm is as the whole population separated into numerous subgroups, firefly permits to find local optimal simultaneously in each community. Therefore, as the population in the network amplifies, there is no effect on the computation time to find local optimum.
- The third advantage of the firefly algorithm is that the control parameter, light absorption coefficient γ can control as the iterations in execution to swift and speed up the chances of converges. That means as the result converges; the procedure discontinues the iterations and locates the optimal value for the control parameters.

5. Experiment and evaluation

In this segment, we applied the proposed algorithm on synthesized and real data sets, and evaluated the algorithm. Initially, we investigated the datasets and next compared the results with the standard SNA measures that are used to find opinion leader in the social network.

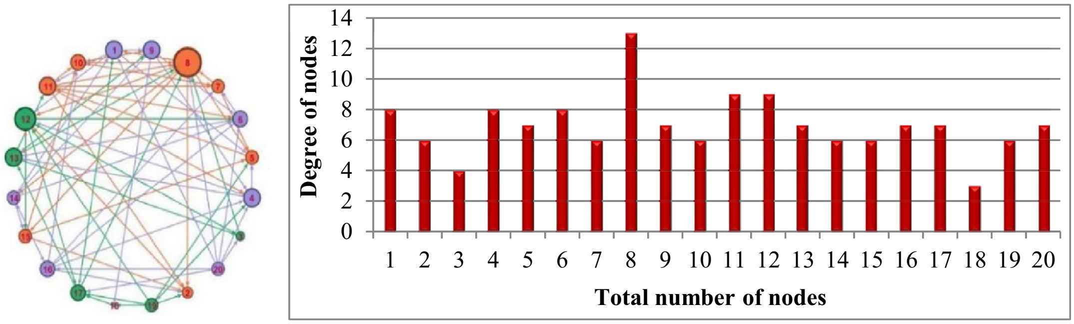
5.1. Dataset

5.1.1. Synthesized dataset

The synthesized dataset has the total of 20 nodes and 70 edges as shown in Fig. 3(a). The density of the network is 7.00, and the degree of each node is illustrated in Fig. 3(b). Now, we have applied a modified Louvain community detection algorithm and found the total four communities for the mentioned dataset. In the next step, we implemented the proposed modified firefly algorithm to compute the attractiveness of each user in each community to find out the local opinion leaders according to their attractiveness. The attractiveness and SNA measures of top-3 users in each community are shown in Table 2. We also applied the same algorithm for the whole network structure to identify the opinion leader at the global level. Once we computed the attractiveness of each user, ranked the user according to their attractiveness and found the top-N (=5) number of global opinion leaders in the network as shown in Table 3.

5.1.2. Real dataset

We implemented our algorithm on a real dataset named 'small slashdot', which is a network of friends and foes (Al-Oufi, Kim, & El Saddik, 2012; Tang, Lou, & Kleinberg, 2012) as shown in Fig. 4. The dataset has total 13,182 nodes as users, 34,621 edges represent the friend relationship among the users, and the density of the network is 5.1981. There are 76.7% users related with friend relationship and rest of the users associated with foes relationship. We have applied the same procedure for this dataset as we had applied on the synthesized dataset.



(a) (b)

Fig. 3. (a) Synthesized social network architecture (b) Node degree distribution.

Table 2
Top-3 Local opinion leader based on various SNA measures and proposed algorithm in each community for the synthesized dataset.

Local community	Node id	DC	Node id	BC	node id	CC	Node id	Eigenvector	Node id	PageRank	Node id	Firefly attractiveness
1	12	0.473684	12	0.063464	12	0.655172	12	0.26998632	12	0.062541	**12**	**0.218250008**
	2	0.315789	19	0.035791	2	0.575758	2	0.198201878	19	0.045186	**2**	**0.119558582**
	15	0.315789	2	0.022086	15	0.575758	15	0.194527396	15	0.043607	**15**	**0.118810812**
2	8	0.684211	8	0.142888	8	0.762192	8	0.377563673	8	0.086868	**8**	**0.392312044**
	1	0.421053	17	0.045134	1	0.633333	1	0.263850871	1	0.055526	**1**	**0.180753303**
	13	0.368421	1	0.034884	20	0.612903	13	0.224548087	17	0.050955	**17**	**0.153157071**
3	11	0.473684	4	0.048933	4	0.633333	11	0.28789483	11	0.061513	**11**	**0.208180898**
	4	0.421053	11	0.042203	11	0.633333	4	0.237325065	4	0.056041	**4**	**0.183708542**
	5	0.368421	16	0.033431	5	0.612903	5	0.229619483	16	0.050074	**5**	**0.151071918**
4	6	0.421053	9	0.057866	6	0.633333	6	0.237956773	6	0.055904	**6**	**0.181841781**
	9	0.368421	6	0.041133	9	0.59375	14	0.194342717	9	0.051139	**9**	**0.152272076**
	14	0.315789	14	0.020969	14	0.575758	9	0.189659725	14	0.043487	**14**	**0.119333331**

Table 3
Top 5 global opinion leader based on various SNA measures and proposed algorithm for the synthesized dataset.

Node id	DC	Node id	BC	Node id	CC	Node id	Eigenvector	Node id	PageRank	Node id	Firefly attractiveness
8	0.684211	8	0.142888	8	0.76	8	0.377564	8	0.086868	**8**	**0.392312044**
12	0.473684	12	0.063464	12	0.655172	11	0.287895	12	0.062541	**12**	**0.218250008**
11	0.473684	9	0.057866	4	0.633333	12	0.269986	11	0.061513	**11**	**0.208180898**
4	0.421053	4	0.048933	11	0.633333	1	0.263851	4	0.056041	**4**	**0.183708542**
6	0.421053	17	0.045134	6	0.633333	6	0.237957	6	0.055904	**6**	**0.181841780**

Fig. 4. Structure of 'small slashdot' social network.

In the real dataset, there is some missing relationship value between the users also exist. Therefore, we use the arithmetic mean of the degrees of the entire node for filling those missing values. Initially, we have measured the total degree of trust for each user for measuring the prestige. Further, we applied the modified Louvain community partitioning algorithm on this dataset and identified the total of 28 community structures. The identification of the community depends upon the type of the network, and the attractiveness of the user relies on the landscape of the network that includes the overall knowledge about the network. The in-degree, out-degree and total degree distribution of the entire nodes of the network, are shown in Fig. 5.

Next, we used the experimental value of all the heuristic parameters for the dataset and applied the modified firefly algorithm. The firefly attractiveness of the top opinion leader of each community is shown in Table 4, and the top-10 global opinion leader identified by our approach and other SNA measures, calculated on the same dataset, are shown in Table 5.

In the mentioned tables, we have compared the results obtained from our algorithm with the Social Network Analysis (SNA) standard measures. Each table includes total six columns, and each column has two sub-columns; the first sub-column contains the node id and second sub-column consists of the value for the particular measure, i.e., in Table 5, the degree centrality of the node having id 936 is 0.031409. Similarly, for both real and synthesized dataset, we have found four tables. Tables 2 and 3 are for the synthesized dataset to discover opinion leader in local and global leader respectively while Tables 4 and 5 are for real dataset to determine opinion leader in local and global leader respectively. The last column in each of the table displays the firefly attractiveness of the node according to the proposed method. Although there are no standard methods for identifying the opinion leader in the social network and it is also very hard-hitting to describe the accuracy and precision of an algorithm.

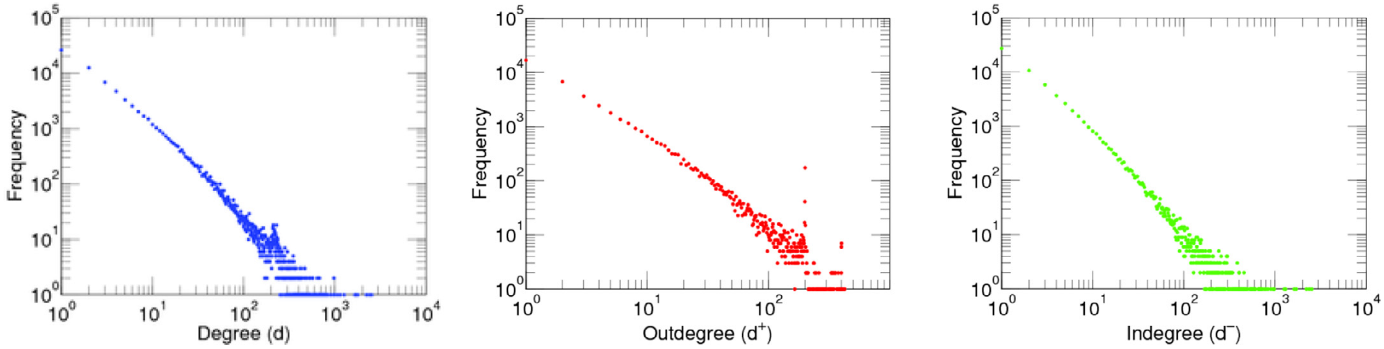


Fig. 5. Degree, Out-degree, and In-degree distribution of network.

Table 4
Top Local opinion leader based on various SNA measures and proposed algorithm in each community for real dataset.

Comm- unity	Node id	DC	Node id	BC	Node id	CC	Node id	Eigenvector	Node id	PageRank	Node id	Firefly attractiveness
1	8	0.025036	8	0.050489	8	0.399739	8	0.114438	8	0.003896	**8**	**0.600012**
2	757	0.013656	757	0.013203	237	0.362853	237	0.015079	757	0.002864	**757**	**0.512457**
3	822	0.020256	822	0.021471	822	0.353719	822	0.019695	822	0.00405	**822**	**0.491772**
4	898	0.021925	898	0.022510	898	0.364136	898	0.03599	898	0.004114	**898**	**0.324512**
5	190	0.030802	190	0.036551	190	0.374471	190	0.079614	190	0.005791	**190**	**0.598897**
6	520	0.011683	520	0.013942	454	0.338434	520	0.005383	520	0.002719	**520**	**0.356568**
7	394	0.017449	394	0.020450	394	0.337542	22	0.009751	394	0.004004	**394**	**0.462458**
8	163	0.016918	163	0.021447	163	0.352867	155	0.019351	163	0.003862	**163**	**0.495554**
9	522	0.029436	522	0.037023	522	0.375666	522	0.051219	522	0.005496	**522**	**0.596478**
10	825	0.029740	825	0.031148	825	0.35607	607	0.06069	825	0.005823	**825**	**0.594271**
11	834	0.014946	834	0.016314	935	0.356407	834	0.018006	834	0.003184	**834**	**0.426680**
12	523	0.018436	523	0.026269	523	0.371432	523	0.020027	523	0.003503	**523**	**0.345609**
13	617	0.016539	905	0.019068	617	0.351397	905	0.028331	905	0.003451	**617**	**0.487541**
14	642	0.030726	642	0.033205	642	0.359018	642	0.030803	642	0.006029	**642**	**0.595370**
15	936	0.031409	936	0.024885	936	0.396266	913	0.166398	791	0.004299	**936**	**0.594602**
16	62	0.031034	62	0.027942	62	0.380075	62	0.125934	62	0.004781	**62**	**0.602541**
17	162	0.018056	162	0.018683	106	0.371149	106	0.02064	162	0.003867	**106**	**0.587458**
18	344	0.013582	344	0.015740	344	0.361629	344	0.032871	344	0.002594	**344**	**0.452112**
19	644	0.018208	644	0.019741	644	0.346112	644	0.013661	644	0.00382	**644**	**0.475556**
20	57	0.017677	57	0.021058	57	0.368772	57	0.042123	669	0.003462	**57**	**0.482331**
21	184	0.031105	184	0.085150	184	0.385669	653	0.086983	184	0.006851	**184**	**0.601245**
22	794	0.015932	794	0.014683	794	0.354271	893	0.046384	794	0.002878	**794**	**0.548854**
23	706	0.000607	706	0.001062	706	0.262063	706	0.000162	706	0.000251	**706**	**0.574845**
24	855	0.011532	855	0.014769	453	0.338365	453	0.005211	855	0.003007	**453**	**0.554002**
25	142	0.020712	142	0.022020	142	0.361837	142	0.033016	142	0.003955	**142**	**0.495822**
26	173	0.025871	173	0.031230	797	0.353928	797	0.035669	173	0.005691	**173**	**0.521406**
27	813	0.015477	813	0.013895	813	0.346131	813	0.036252	813	0.002878	**813**	**0.560002**
28	248	0.011228	248	0.012686	248	0.344107	1072	0.00794	568	0.002602	**248**	**0.485690**

Table 5
Top-10 Global opinion leader based on various SNA measures and proposed algorithm for real dataset.

Node id	DC	Node id	BC	Node id	CC	Node id	Eigenvector	Node id	PageRank	Node id	Firefly attractiveness
936	0.031409	184	0.085150	8	0.399739	913	0.166398	184	0.006851	**62**	**0.602541**
184	0.031105	8	0.050489	936	0.396266	62	0.125934	642	0.006029	**184**	**0.601245**
62	0.031030	522	0.037023	82	0.390757	8	0.114438	825	0.005823	**8**	**0.600012**
190	0.030802	190	0.036551	43	0.388648	653	0.086983	190	0.005791	**190**	**0.598897**
642	0.030726	642	0.033205	625	0.387688	190	0.079614	173	0.005690	**522**	**0.596478**
913	0.030726	173	0.031230	913	0.385872	607	0.06069	522	0.005496	**642**	**0.595370**
82	0.030574	825	0.031148	184	0.385669	522	0.051219	62	0.004780	**913**	**0.594829**
791	0.030498	62	0.027942	791	0.383302	893	0.0463	791	0.004299	**936**	**0.594602**
825	0.029740	523	0.026269	74	0.381494	57	0.042123	898	0.004114	**825**	**0.594271**
522	0.029436	936	0.024885	62	0.380075	813	0.036252	822	0.004050	**791**	**0.594158**

Further, in our investigation, we identified that there is a correlation between the light absorption coefficient γ and the size of the network, we analyzed that changes in the values of light absorption coefficient γ and the size of the network, also affect the correlation coefficient r. As soon as we modify the value of the light absorption coefficient γ, the amount of correlation coefficient also changes as shown in Fig. 6.

We can infer that as the light absorption coefficient increase initially, correlation coefficient also increases but later it decreases gradually. Hence, we found that the correlation coefficient r also changes as the number of opinion leader changes as shown in Table 6.

Furthermore, the experimental result also indicates that if we compared the findings proposed by our algorithm to the other methods used for the same, the total number of opinion leaders identified in each community is also varied. Now, we have demonstrated top 5% opinion leaders out of the total number of users dis-

Table 6
Correlation coefficient r and Light absorption coefficient γ relationship.

Light absorption coefficient (γ)	$r(N=10)$	$r(N=100)$	$r(N=500)$	$r(N=1000)$	$r(N=5000)$	$r(N=10{,}000)$
0.2	0.3250	0.3255	0.4021	0.4350	0.4822	0.5214
0.5	0.3511	0.3758	0.4832	0.5298	0.6235	0.7566
0.7	0.3355	0.3473	0.4521	0.4752	0.5214	0.6323

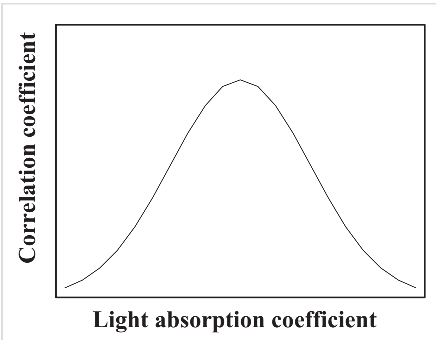


Fig. 6. Absorption and Correlation coefficient relation.

covered by each method in the social network for the real dataset as shown in Fig. 7.

We can observe that the proposed algorithm found only those users as opinion leaders who are highly competent and deserving in the communities.

For the validation of the proposed method, we compared the results produced by the proposed research with other SNA measures concerning the accuracy, precision, recall, and F1-score (POWERS, 2011). For measuring all the performance metrics, we exploit the ground truth of the communities. Ground truth provides the actual information about communities, and the total number of opinion leaders exists in those communities. There are four performance metrics: Accuracy, Precision, Recall, and F1-score that indicate how good our prototype is (Makhoul, Kubala, Schwartz, & Weischedel, 1999; Thagard, 2008). There are four parameters: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative are used for measuring these statistical performance metrics (François, 2006). Let us discuss each parameter in brief as follow:

True Positive (TP): These values are the valid projected positive values, i.e., both the actual class and the anticipated class states yes.

True Negative (TN): These values are the valid projected negative values, i.e., both the actual class and the anticipated class states no.

False Positive (FP): These values are the false projected positive values, i.e., the actual class states no and the anticipated class states yes.

False Negative (FN): These values are the false projected negative values, i.e., the actual class states yes and the anticipated class states no.

Now, we briefly discuss the Accuracy, Precision, Recall, and F1-score as follow:

Accuracy: Accuracy of the model can be defined as the ratio between total correctly predicted value and the total no of predicted values. It is the most significant factor to evaluate that how best our proposed model is.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision: Precision of the model can be defined as the ratio between the acceptably forecasted positive values and total positive values. If the precision is very high, it assured that the model finds the lesser number of false positive values.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall: Recall of the model can be defined as the fraction between the acceptably forecasted positive values and all the yes observations in actual class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score: F1-score is the subjective average of recall and precision. Therefore, it takes both false positive and false negative into account.

$$\text{F1} - \text{score} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

For example, consider a social network; in which total 100 nodes are known opinion leader from provided ground truth. According to the proposed model, if 73 nodes found as true positive, 12 nodes found as true negative, five nodes found as false positive and ten nodes found as a false negative, the Accuracy, Precision, Recall, and F1-score of the model is 0.8500, 0.935, 0.879, and 0.906 respectively. We can infer that our approach produces better results as compared to other standard measures as shown in Fig. 8.

Additionally, we also observed that when the values of the heuristic control parameters revolutionize in a different province, the total amount of dignified opinion leader might vary. The crux of the research is to optimize the attractiveness of the user that persuades other assessment and perception about a particular object. Firefly algorithm is appropriate for the social network because the behavior of the firefly matched with the user's behavior of the social network. Moreover, we also observed that the total computation time taken by our algorithm is also very less as compared to other SNA measure as shown in Fig. 9.

The modified Louvain method is helpful to uncover the essential communities. We have compared the average running time and modularity value between the original and the modified Louvain method as shown in Table 7 and found the improved and optimized result. Therefore, confidently we can say that proposed algorithms produce better outcomes as compared to other standard SNA measures.

6. Strength and weakness of the proposed method

As we have discussed in the previous section that until now, nearly no research work has been done using nature-inspired algorithms to identify opinion leader. Although, over the earlier period, many researchers proposed various approaches to find out the opinion leader in the social network but none of them had suggested any heuristic move toward the same. Nature inspired algorithm is the excellent resource of stimulation for cracking real world tribulations. The first time, we have introduced the nature-inspired firefly algorithm to find out the opinion leaders in the social network. These techniques are too expandable and resilience to resolve any complex problem as the size and aptitude of the

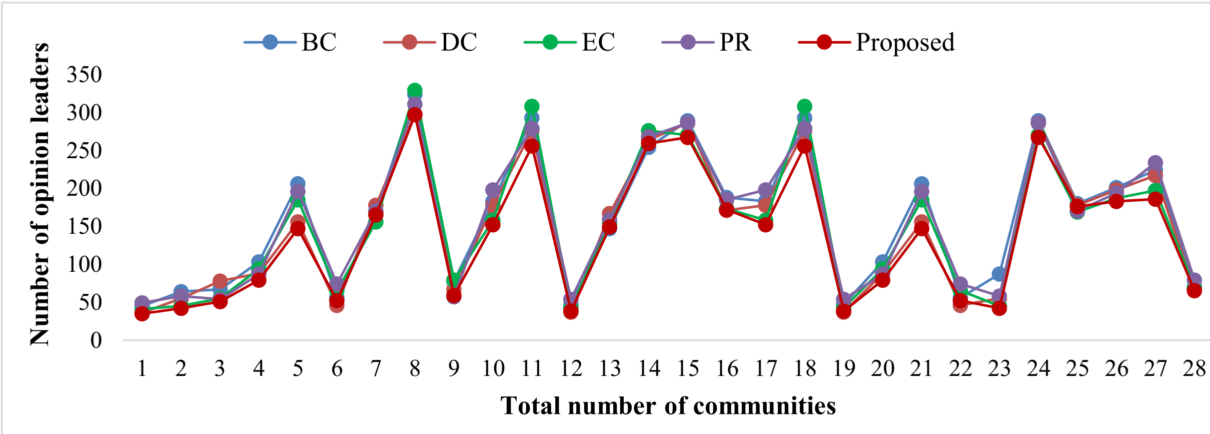


Fig. 7. Comparison of BC, DC, EC, PR, and proposed approach for the top 5% users in each community for real data set.

Table 7
Comparison between proposed and original Louvain method for real and synthesized data set.

Attributes	Original Louvain		Proposed Louvain	
	Real	Synthesized	Real	Synthesized
No of nodes	13,182	20	13,182	20
Modularity	0.8522	0.542	0.7237	0.483
No of runs	100	10	100	10
Average running time (in sec.)	749	18	638	11

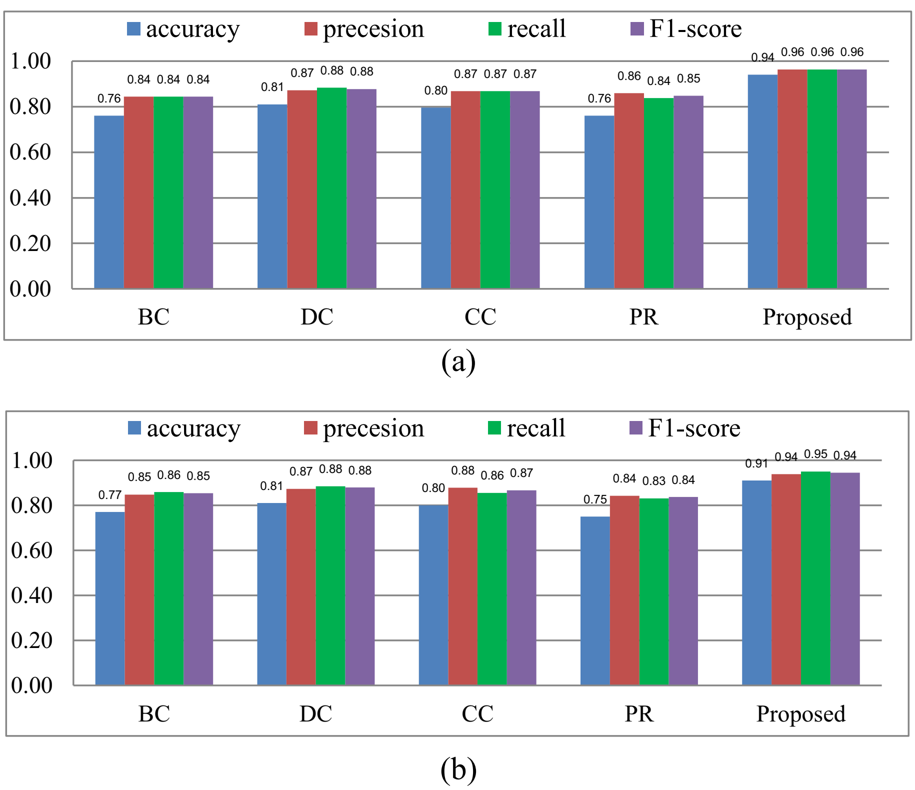


Fig. 8. Comparison of accuracy, precision, recall, and F_1-score for (a) real data set (b) synthesized data set.

problem increases (Fister et al., 2013). The proposed method can discover the opinion leader more accurately and precisely even though the network size is increasing. For legalize the above statement, we have compared the results with the other standard SNA measures and found that our proposed research work is much better in term of accuracy, precision, F1-score, recall and computation time. The proposed method behaves like an intelligent system because each user computes its prominence value repeatedly from various centrality measures whenever the user's position changes or a new node added in the network. It wisely discovers only those

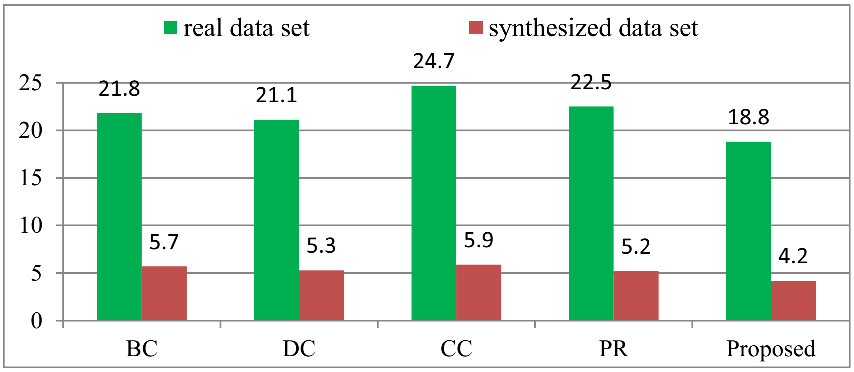


Fig. 9. Comparison of computation time for the real and synthesized data set.

users who are competent to be an opinion leader. No other previously developed algorithm was able to handle this feature. The previous researchers were able to find the opinion leader only within the social network, but the proposed model discovers the opinion leader not only on the global level but on the local level within each community also. Therefore, it provides inter-network and intra-network opinion leader. The proposed model also includes the trust factor for identifying the light absorption coefficient γ of each user. The light absorption coefficient γ is primarily used to measure the prestige that eventually measures the prominence of the user. Hence, the proposed research work also includes the trust factor as well. Besides, the modified Louvain method is intended to discover the communities in the social network build on the modularity gain of the network. The proposed Louvain method also includes the clustering coefficient to attain modularity gain.

The weakness of the proposed model is that it can discover the opinion leader in the static social network. In the social network, the relationship between the users and the number of users in the network changes with time. In the future, we would make an effort to propose the same algorithm for the dynamic social network too. In the social network, there exhibit lots of copious features such as total number of re-tweets, text description, number of followers, user's sentiment, the total number of user's login, user's interest, and many more. Our proposed model only includes trust, clustering coefficient, the distance between the user, and centrality to measure the attractiveness of each user. In the future, we would attempt to include some more features to measure attractiveness, so that we could find opinion leader more precisely. There are various other nature-inspired algorithms exists, such as PSO, Bee optimization, Ant colony optimization and many more but we have considered only the firefly algorithm in this research. In the future, we would also make an effort to find out the opinion leader using other nature-inspired algorithms.

7. Conclusion and future scope

We have introduced a novel approach to discover the local and global opinion leader in the social network communities using heuristic firefly search algorithm. Although, it is very tedious to find out the competent, deserving and knowledgeable opinion leader in social network, yet we have attempted to solve the problem up to a certain level. There are various other factors such as trust (Aghdam & Jafari Navimipour, 2016; Dorigo & Di Caro, 1999), textual content (Dubois & Gaffney, 2014), user's opinion and sentiment (Duan et al., 2014) and many other factors that we discussed earlier, can also be associated with our algorithm to discover the opinion leader more precisely.

In our proposed algorithm, initial, we identified the communities using the modified Louvain community partitioning algorithms in which the concept of clustering coefficient is associated to find out the communities. Next, we found the local and global opinion leader using the firefly algorithm that produces the better result as compared to other SNA measures, and the results indicate that the proposed algorithm finds the optimal opinion leaders.

Although the proposed approach can be implemented on lots of real-world applications but due to space constraints, we are considering only a few applications. One of the primary real-world applications of this model is a recommendation system that recommends the implications based on user interest and importance. If the person's earlier period activities are parallel to some other person's actions, in the future, it may be possible that the same person also resembles the same actions. The proposed model is very conducive to recommend the best products, reviews, friendship, movies, hotels, and so on with the help of opinion leader in that particular domain. In this application, the user will behave as a firefly, and the entire users will search for another user having more expertise in a specific field. That user will consider as an opinion leader who provides the recommendation to other users. The traditional recommended system pays no attention to the relationship among the entire users, but in the social network relationship among the users present that further may be helpful for critical decisions. Therefore, the proposed model not only obliging the optimal suggestions but also considers the relationships among the user in the social network. The proposed model can also be regarded as an expert system if it recommends the activities and actions using Artificial Intelligence (Katarya & Verma, 2016a, 2016b).

The proposed model will be cooperative for disaster management that provides the architecture to handle the natural disaster such as earthquake, volcanoes, flood, drought and many more by their decision-making capabilities using a social network and AI techniques (Underwood, 2010). The developed model will be very beneficial and assistive if it is intelligently figured out the proper resource management among the areas, which are deeply affected by natural disaster. In this application, whenever any natural disaster occurs in any area, the user will share the information through the social networking sites. The model will use their intelligence and take administrative decisions to reduce economic misplace as much as possible.

In the future, the proposed model can be used as an expert system to amalgamate all components of agriculture such as water management, soil management, disease management, crop management, rainwater management, resource management and many more into a scaffold which concentrate on the best optimal possible solution for the farmers to get rid of any agriculture-related problem (Zhang, Wang, & Wang, 2002). In this framework, the entire farmers will consider as firefly that updates their attractiveness values by searching the more competent, knowledgeable, and skilled farmers using the rule-based and knowledge base technique incorporate with a social network. This framework is more suitable in the current scenario because still most of the farmers are not aware of modern agricultural tools, methods, and techniques. With the help of this model, we will be able to find out the knowledgeable opinion leader in the agriculture domain.

Another application of the proposed model is in crime management. In this approach, we can identify the most notorious and dangerous person in the set of the criminal person by considering the centrality of the node, clique, and k-core element in the network. Hence, in this application, the most notorious person will be viewed as a leader against which most of the criminal cases filled, and these people are the core leader of the group. We can also track the person's scandalous background by sharing the information with other cops and update their records accordingly. We can also discover the most cohesive set of criminals via this approach. Therefore, In the future, the proposed model will be conducive to identify the criminals and in some cases, can reduce the crime rate too.

The proposed model can also be used to discover optimal researchers in the educational domain. In the educational realm, the lot of researchers exists, but it is very challenging to identify the most active researchers in a specialized area. Hence, the proposed approach can behave as an expert system that will classify the researchers into different categories using other researcher's feedback incorporated with AI techniques.

Furthermore, in the future, we can also use the other meta-heuristic and nature-inspired algorithms such as PSO, Cuckoo search, Whale optimization, and many others to discover the opinion leader (Li, Ma, Zhang, & Huang, 2013; Luo et al., 2018; Yang, 2010). Even in the future, we will also concern to improve the accuracy of the algorithm by considering the network topology, user's opinion, and user's tweets.

One of the limitations of this algorithm is that it implemented on the static social network. Now a day, networks changes rapidly and the relationship between the users appear and disappear very quickly. Hence, it is required to implement this algorithm on the dynamic social network that indicates the dynamic nature of the network (Skyrms & Pemantle, 2000).

Authors contribution

Following authors have contributed in the manuscript

1 Conceived and designed the analysis
2 Collected the data
3 Contributed data or analysis tools
4 Performed analysis
5 Wrote the paper
6 All other necessary contributions related to this paper

References

Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: a computer simulation. *Organization Science*. doi:10.1287/orsc.8.3.289.

Adali, S., Escriva, R., Goldberg, M. K., Hayvanovych, M., Magdon-Ismail, M., Szymanski, B. K., et al. (2010). Measuring behavioral trust in social networks. *ISI 2010 - 2010 IEEE international conference on intelligence and security informatics*. Public Safety and Security. doi:10.1109/ISI.2010.5484757.

Aghdam, S. M., & Jafari Navimipour, N. (2016). Opinion leaders selection in the social networks based on trust relationships propagation. *Karbala International Journal of Modern Science, 2*(2), 88–97. doi:10.1016/j.kijoms.2016.02.002.

Al-Oufi, S., Kim, H. N., & El Saddik, A. (2012). A group trust metric for identifying people of trust in online social networks. *Expert Systems with Applications*. doi:10.1016/j.eswa.2012.05.084.

Aleahmad, A., Karisani, P., Rahgozar, M., & Oroumchian, F. (2016). OLFinder: Finding opinion leaders in online social networks. *Journal of Information Science*. doi:10.1177/0165551515605217.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. WWW 2012 - Session: Information Diffusion in Social Networks April 16-20, 2012, Lyon, France. doi: 10.1145/2187836.2187907

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports*. doi:10.1016/j.physrep.2005.10.009.

Bodendorf, F., & Kaiser, C. (2010). Detecting opinion leaders and trends in online communities. In *4th international conference on digital society, ICDS 2010, includes CYBERLAWS 2010: The 1st International conference on technical and legal aspects of the e-society* (pp. 124–129). doi:10.1109/ICDS.2010.29.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*. doi:10.1080/0022250X.1972.9989806.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*. doi:10.1111/j.1083-6101.2007.00393.x.

Chan, K. K., & Misra, S. (1990). Characteristics of the opinion leader: A new dimension. *Journal of Advertising*. doi:10.1080/00913367.1990.10673192.

Chen, Y., Hui, L., Wu, C. I., Liu, H., & Chen, S. (2017). Opinion leaders discovery in dynamic social network.

Cho, Y., wang, J., & Lee, D. (2012). Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach. *Technological Forecasting and Social Change, 79*(1), 97–106. doi:10.1016/j.techfore.2011.06.003.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. Cond-Mat/0408187. doi: Doi: 10.1103/PhysRevE.70.066111

Dichter, E. (1966). How Word of mouth advertising works. *Harvard Business Review*. doi:10.1016/S0267-3649(00)88914-1.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*. doi:10.1145/502512.502525.

Dorigo, M., & Di Caro, G. (1999). Ant colony optimization: A new meta-heuristic. In *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*. doi:10.1109/CEC.1999.782657.

Duan, J., Zeng, J., & Luo, B. (2014). Identification of opinion leaders based on user clustering and sentiment analysis. In *Proceedings - 2014 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology - workshops* WI-IAT 2014. doi:10.1109/WI-IAT.2014.59.

Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: identifying political influentials and opinion leaders on twitter. *American Behavioral Scientist*. doi:10.1177/0002764214527088.

Dye, R. (2000). The buzz on buzz. *Harvard Business Review*. doi:10.1097/00043764-200106000-00013.

Fister, I., Yang, X. S., & Brest, J. (2013). A comprehensive review of firefly algorithms. *Swarm and Evolutionary Computation*. doi:10.1016/j.swevo.2013.06.001.

François, D. (2006). Binary classification performances measure cheat sheet. *Journal of Machine Learning Research*. doi:10.1109/TAP.2013.2272673.

Freeman, L. C. (1978a). Centrality in social networks. *Social Networks*. doi:10.1016/0378-8733(78)90021-7.

Freeman, L. C. (1978b). Centrality in social networks conceptual clarification. *Social Networks*. doi:10.1016/0378-8733(78)90021-7.

Gold, D., Katz, E., Lazarsfeld, P. F., & Roper, E. (1956). Personal influence: The part played by people in the flow of mass communications. *American Sociological Review*. doi:10.2307/2088435.

Goyal, A., Bonchi, F., & Lakshmanan, L. (2008). Discovering leaders from community actions. In *Proceedings of the 17th ACM conference on information and knowledge management*. doi:10.1145/1458082.1458149.

Heidemann, J., Klier, M., & Probst, F. (2010). Identifying key users in Online Social Networks: A PageRank based approach. In *Proceedings of the 31st international conference on information systems*. doi:10.1016/j.infsof.2008.09.005.

Katarya, R., & Verma, O. P. (2016a). A collaborative recommender system enhanced with particle swarm optimization technique. *Multimedia Tools and Applications*. doi:10.1007/s11042-016-3481-4.

Katarya, R., & Verma, O. P. (2016b). Recent developments in affective recommender systems. *Physica A: Statistical Mechanics and Its Applications*. doi:10.1016/j.physa.2016.05.046.

Li, F., & Du, T. C. (2011). Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs. *Decision Support Systems, 51*(1), 190–197. doi:10.1016/j.dss.2010.12.007.

Li, Y., Ma, S., Zhang, Y., Huang, R., & Kinshuk (2013). An improved mix framework for opinion leader identification in online learning communities. *Knowledge-Based Systems, 43*, 43–51. doi:10.1016/j.knosys.2013.01.005.

Liu, C., Tian, Y., Zhang, Q., Yuan, J., & Xue, B. (2013). Adaptive firefly optimization algorithm based on stochastic inertia weight. In *Proceedings - 6th international symposium on computational intelligence and design* ISCID 2013. doi:10.1109/ISCID.2013.90.

Luo, L., Yang, Y., Chen, Z., & Wei, Y. (2018). Identifying opinion leaders with improved weighted LeaderRank in online learning communities. *International Journal of Performability Engineering*. doi:10.23940/ijpe.18.02.p1.193201.

Ma, N., & Liu, Y. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. *Expert Systems with Applications*. doi:10.1016/j.eswa.2013.08.033.

Mak, V. (2008). The emergence of opinion leaders in social networks. *SSRN Electronic Journal*. doi:10.2139/ssrn.1157285.

Makhoul, J., Kubala, F., Schwartz, R., & Weischedel, R. (1999). Performance measures for information extraction. In *Broadcast news workshop '99 Proceedings*. doi:10.1016/j.cell.2008.03.030.

Myers, J. H., & Robertson, T. S. (1972). Dimensions of opinion leadership. *Journal of Marketing Research*. doi:10.2307/3149604.

Newman, M. E. J. (2006). Modularity and community structure in networks. In *Proceedings of the national academy of sciences*. doi:10.1073/pnas.0601602103.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. doi:10.1016/j.socnet.2010.03.006.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. *World Wide Web Internet And Web Information Systems* doi: 10.1.1.31.1768.

POWERS, D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies* doi: 10.1.1.214.9232.

Ruhnau, B. (2000). Eigenvector-centrality - a node-centrality. *Social Networks*. doi:10.1016/S0378-8733(00)00031-9.

Scott, J. (2000). *Social network analysis: a handbook*. SAGE Publications.

Shafiq, M. Z., Ilyas, M. U., Liu, A. X., & Radha, H. (2013). Identifying leaders and followers in online social networks. *IEEE Journal on Selected Areas in Communications*. doi:10.1109/JSAC.2013.SUP.0513054.

Sharara, H., Getoor, L., & Norton, M. (2011). Active surveying: A probabilistic approach for identifying key opinion leaders. *Twenty-second international joint conference on artificial intelligence*.

Sherchan, W., Nepal, S., & Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys*. doi:10.1145/2501654.2501661.

Skyrms, B., & Pemantle, R. (2000). A dynamic model of social network formation. In *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.97.16.9340.

Song, X., Chi, Y., Hino, K., & Tseng, B. (2007). Identifying opinion leaders in the blogosphere. In *Proceedings of the sixteenth ACM conference on conference on information and knowledge management - CIKM '07* (p. 971). doi:10.1145/1321440.1321588.

Sthle, L., & Wold, S. (1989). Analysis of variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*. doi:10.1016/0169-7439(89)80095-4.

Tang, J., Lou, T., & Kleinberg, J. (2012). Inferring social ties across heterogenous networks. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. doi:10.1145/2124295.2124382.

Thagard, P. (2008). Machine learning. *In A Companion to Cognitive Science*. doi:10.1002/9781405164535.ch16.

Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. *Journal of Marketing Research*. doi:10.1509/jmkr.47.4.643.

Underwood, S. (2010). Improving disaster management. *Communications of the ACM*. doi:10.1145/1646353.1646362.

Van der Merwe, R. &, & Van Heerden, G. (2009). Finding and utilizing opinion leaders: Social networks and the power of relationships. *Outh African Journal of Business Management*.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. *Structural Analysis in the Social Sciences, 8*. doi:10.1525/ae.1997.24.1.219.

Yang, X.-S. (2010). Nature-inspired metaheuristic algorithms. *Nature-Inspired Metaheuristic Algorithms Second Edition*. doi:10.1016/B978-0-12-416743-8.00005-1.

Yang, X. S., & He, X. (2013). Firefly algorithm: Recent advances and applications. *International Journal of Swarm Intelligence*. doi:10.1504/IJSI.2013.055801.

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*. doi:10.1016/S0168-1699(02)00096-0.

Zhao, Y., Kou, G., Peng, Y., & Chen, Y. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences*. doi:10.1016/j.ins.2017.10.031.

Zhu, M., Lin, X., Lu, T., & Wang, H. (2016). Identification of opinion leaders in social networks based on sentiment analysis : Evidence from an automotive forum. *Advances in Computer Science Research*.