



Opinion Leader Detection in Online Social Networks Based on Output and Input Links

Zahra Ghorbani¹ · Saeid Ghafouri¹ · Seyed Hossein Khasteh¹

Accepted: 28 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The understanding of how users in a network update their opinions based on their neighbours' opinions has attracted a great deal of interest in the field of network science, and a growing body of literature recognises the significance of this issue. In this work, we propose a new dynamic model of opinion formation in directed networks. In this model, the opinion of each node is updated as the weighted average of its neighbours' opinions, where the weights represent social influence. We define a new centrality measure as a social influence metric based on both influence and conformity. We measure this new approach using two opinion formation models: (i) the Degroot model and (ii) our own proposed model. Previously studies have not considered conformity, and have only considered the influence of the nodes when computing the social influence. In our definition, nodes with low in-degree and high out-degree that were connected to nodes with high out-degree and low in-degree had higher centrality. As the main contribution of this research, we propose an algorithm for finding a small subset of nodes in a social network that can have a significant impact on the opinions of other nodes. Experiments on real-world data demonstrate that the proposed algorithm significantly outperforms previously published state-of-the-art methods.

Keywords Social networks · Social influence · Opinion formation · Identification of influential nodes · Degroot model

Zahra Ghorbani and Saeid Ghafouri have contributed equally to this work.

✉ Saeid Ghafouri
sdghafouri@gmail.com

Zahra Ghorbani
z.ghorbani@mail.kntu.ac.ir

Seyed Hossein Khasteh
khasteh@kntu.ac.ir

¹ Department of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

1 Introduction

Recently, a large number of research studies have been undertaken to achieve a better understanding of individuals' everyday opinions on various subjects. This increase in attention is due to the fact that opinions crucially shape human behaviour in important areas such as economics, political science, cultural problems etc. Social networks play a fundamental role in the forming and reshaping of public opinions, and have therefore attracted a great amount of attention in the area of opinion formation studies [11, 28, 37, 39, 45]. Consider a prominent application in a viral marketing campaign that aims to convert a small number of initial users in order to eventually influence the majority of people in the market. This problem can be viewed as an example of finding "opinion leaders" in a social network, which is also the main challenge of this research paper. A further example might involve the forecasting of the final opinion of each node and the maximisation of the spread of influence via a network [9, 23, 38]. Society can also be seen in terms of a network, and we can therefore apply our approach for finding opinion leaders to a large number of social problems such as control of the spread of disease [59, 62], game theory based on financial problems [44, 52] and any other problems related to finding the influential nodes in an arbitrary network [2, 10, 18, 33, 48, 58, 60].

We should emphasise that our method should not be confused with influence maximisation research studies that operate under certain propagation models [22, 34, 35, 61]. Our research paper finds opinion leaders using opinion formation models, which despite certain similarities has critical differences from influence maximisation studies. Firstly, influence maximisation studies are based on the solution of an optimisation problem through leveraging some kind of greedy algorithm, and although the details of these vary from one research study to another, they all use the same technique, as mentioned above. However, our research paper takes a different approach involving the use of an iterative algorithm. This paper aims to find the values of a concept called "centrality" for each node in order to use these further in finding opinion leaders. Centrality is a measure of the relative importance of each node in a network based on its connectivity or influence within the network. Centrality metrics vary in how they define and quantify importance, common metrics include degree centrality, which counts the number of direct connections a node has, and betweenness centrality, which quantifies how often a node serves as a bridge along the shortest path between two other nodes. These centrality measures help identify nodes that are potentially influential or central to the network's structure and dynamics. In addition, we test opinion formation models. In most influence maximisation studies, each node has two Boolean values. A node is either active or inactive, and unlike opinion formation models, does not have continuous values. We explain our opinion formation model in more detail in Sect. 2.3.

The main approaches to opinion formation theory can be divided into methods based on (i) *Bayesian learning* and (ii) *non-Bayesian learning*. Opinion formation based on Bayesian updates, such as the work presented in [8, 24, 30, 43] involves updating the opinion of each user based on the Bayes theorem. The difficulty with Bayesian methods is that they generally form and update opinions based on the underlying state of and signals from the world, and consequently the results obtained by such methods cannot be easily generalised to other states of the network. This does not mean that these methods are not at all generalizable; however, compared to non-Bayesian methods, they are much less generalizable. Opinion formation methods based on non-Bayesian approaches employ rule-of-thumb methods to form the opinion of agents, giving good approximations of each agent's

behaviour and avoiding the difficulties of Bayesian approaches. These approaches are widely applied in the literature, such as the research studies presented in [6, 11, 17, 32, 51, 56]; these are closely related to the Degroot model [15], and form the foundations of our proposed opinion formation model. Previous researchers in the area of opinion formation studies have typically focused on undirected networks, and have assumed no significant distinction between ingoing and outgoing links in the network. In this paper, we introduce a new model for directed networks in which are two type of links for each node: (i) ingoing; and (ii) outgoing. Outgoing links represent connections between that node and its followers, while ingoing links are the connections to its followers.¹ In our model, the opinion of a node is updated based on the weighted sum of the opinions of its neighbour nodes; the weights of these neighbour nodes are given by their centrality, which we will define in Sect. 2. The main component of this research is based on the concept of centrality, and we design an efficient algorithm for finding the most influential nodes, that is, nodes that impose their own opinion on a significant proportion of the other nodes. Previous centrality models, apart from traditional models such as betweenness or degree centrality, do not consider both ingoing and outgoing links, and thus are not optimally suitable for finding opinion leaders. For example, game centrality aims to find the most suitable nodes to defect in a network modelled with game theory, and can be modelled by a game [52]. Further examples of game theory-based approaches are Shapley centrality [53]. Other works leverage centrality, with the main focus of specific domain of brain networks [31, 47], and perturbation centrality, which is a model inspired by vessel communications, Another work [13] introduces a centrality measure combining return random walk and effective distance gravity model, effectively identifying influential nodes by considering both local, global, and dynamic information. Another example is [1] which examine centrality measures in Protein-Protein Interaction (PPI) networks, focusing on topological and gene annotation approaches to identify essential proteins. Some works have even expanded their criteria to multilayer networks or applications of centrality in aspects of daily life such as urban transport [40]. There are also comprehensive surveys [4, 14, 57] dedicated to introducing the most recent works in the area of centrality. All of these methods perform well within their specific domains or other networks. However, they are not conceptually suitable for finding opinion leaders, this is true even for novel models like k-shell centrality, which shows very good great performance in various domains [36]. In addition, the centrality measures used in other papers to find opinion leaders are specifically designed for the opinion formation networks (e.g. [8, 16, 17, 25–27]). We are not the first to consider opinion networks in our centrality model, since previous works such as [12] consider opinion networks in their centrality models, but we choose a different approach to do so. In all the networks considered here, the experimental results indicate that our proposed method significantly outperforms the other algorithms, such as those based on eigenvector, betweenness and closeness centralities. Section 2 of this paper will discuss methods and materials, while Sects. 3 and 4 present the experimental results and the conclusion and future research respectively.

¹ Other nodes which follow that particular node.

2 Why Another Model?

As mentioned in the introduction section, there are already a number of state-of-the-art methods dedicated to finding influential nodes and opinion leaders, in which more central nodes are referred to under the name of “centrality”. Despite the strong performance achieved by most of these methods, we believe that there are a number of deficiencies that can provide good motivation for seeking a new novel method. Firstly, although there are plenty of research studies dedicated to the problem of finding opinion leaders, their numbers are still lower than the number of articles about influence maximisation. Another reason is that some of the existing research studies are domain-specific, and although their methods show good performance within these specific domains, their performance in other areas such as opinion formation is not guaranteed. A key consideration in our research is considering conformity in our opinion leader detection algorithm. Conformity, in the context of this research, refers to a node’s tendency to align its opinions or behaviours with those prevalent among its neighbouring nodes in a social network. Unlike influence, which measures the capability of a node to affect others, conformity assesses how strongly a node’s opinions are influenced by its peers. This metric considers factors such as the number of incoming links from other nodes, indicating the extent to which a node integrates opinions from its immediate network environment.

From a technical point of view, we believe that our method has the following advantages: (i) conformity and input and output links are considered in our computations; (ii) a heuristic is proposed to reduce the number of computations in each step; and (iii) a novel opinion formation model is proposed that takes the conformity of nodes into account.

Table 1 shows a comparison of several state-of-the-art methods with our proposed method in terms of three technical criteria.

3 Methods and Materials

The main goal of this paper is to solve the problem of measuring the social influence of users in online social networks; this in turn is used as a metric to find the nodes. In real-world problems, the term “social influence” could refer to individuals’ income, persuasiveness of speech, level of education or even their physical attributes, their power among other society ranking members, or the place of each individual in such rankings. Most of the existing research studies [7, 17, 25, 27, 41], measure social influence by raising the question of the extent to which a node can affect other nodes of the network. We extend these methods by taking into account

Table 1 Comparison between different methods

Advantage Method	(1)	(2)	(3)
Our proposed method	✓	✓	✓
[27]	✗	✓	✗
[26]	✗	✓	✗
[58]	✗	✗	✗
[50]	✗	✓	✗
[3]	✗	✓	✗
[8]	✓	✗	✓

another attribute of nodes called conformity. We consider conformity to be a metric of the bias of a node's opinion based on the opinions of the other nodes in the network. As a result, our approach improves upon previous methods by not only considering a node's influence but also using its conformity to evaluate the extent to which a node adheres to its current opinions. For example, consider the Instagram network, in which ingoing links represent the connections to the users that are followed, and outgoing links represent the connections from the user's followers. In citation networks of scientific papers, the ingoing links represent connections among papers that are cited by a given paper, while the outgoing links represent connections to papers that have cited that paper. The conformity of each node is based on a number of parameters, such as the number of its input links, the stability of each node's opinion, the nature of the specific relations in the social network and so forth. All of these parameters should be considered in calculating the conformity of each node; however, to simplify our formulation, we simply omit other factors and consider only the number of input links as a measure of conformity.

3.1 Existing Metrics

The discovery of influential nodes is a challenging issue in the field of complex network analysis. A number of large cross-sectional studies suggest that heuristic methods can efficiently recognise the most influential nodes of networks. The influence of a node is primarily associated with the mathematical definition of a quantity called "centrality", which refers to several measures that define each node's importance in a graph. As shown in the next section, we compare our proposed method for identifying influential nodes using several other measures of centrality such as degree, betweenness, closeness and eigenvector centralities. The simplest measure of centrality is degree centrality, which denotes the number of immediate neighbours of a node. Degree centrality can be interpreted as the number of walks of length one, starting from that particular node, and this measures the local influence of a node. The degree centrality for a directed graph has one of two forms: either the in-degree, which counts the number of direct ties to the node, or the out-degree, which counts the number of ties that a node directs to others. Each of these forms has a different meaning in networks. For comparison purposes, we use out-degree centrality here, in which a node with high out-degree indicates an influential node.

The betweenness centrality [20] is defined as a fraction of the shortest paths between each pair of nodes passing through the considered node. Equation 1 shows the betweenness centrality of node i :

$$g_i = \sum_{j \neq k \neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \quad (1)$$

where σ_{jk} refers to the total number of shortest paths from node j to node k , and $\sigma_{jk}(i)$ denotes the number of these paths passing through i . Another measure of centrality is closeness centrality, defined as the inverse of the average of the shortest-path distance to other nodes. In this measure, nodes with high closeness reach others quickly and play an important role in spreading information. Equation 2 reveals the closeness centrality for a given node i in a directed network as:

$$C_i = \frac{N-1}{\sum_{j=1}^N d(i,j)} \quad (2)$$

Eigenvector centrality, as introduced by Bonacich [5], links nodes to other important nodes, known as central nodes. It defines the centrality of a node based on the centrality of the other nodes connected to it. For a node i with neighbouring nodes $N(i)$, the centrality score can be calculated as follows:

$$c_i = \frac{1}{\lambda} \sum_{j \in N(i)} c_j = \sum_j A_{ij} c_j \quad (3)$$

or in matrix format,

$$AC = \lambda C \quad (4)$$

In general, there are many different eigenvalues where λ refers to a non-zero eigenvector solution. Each row in C represents a node's eigenvector centrality.

Another centrality that is often used for comparison is the PageRank centrality, which is a variant of eigenvector centrality. It was introduced for use by the Google web search engine to rank web pages in the World Wide Web [46]. It ranks web pages based on the sum of the ranks, and can be computed through the following equation:

$$PR(i) = (1 - d) + d \sum_{j \in A_{i,in}} \frac{PR(j)}{d_j^{out}} \quad (5)$$

where N is the total number of pages on the Web, d is a damping factor and d_j^{out} is the out-degree of j . PageRank centrality is currently used to rank nodes in various types of directed networks (e.g. its usage in designing an opinion formation model) [17].

3.2 A New Centrality Measure: Global Centrality

Centrality is a well-known measure in network theory and has several variations in literature reviews [4]. Centrality is defined as the main indicator of the extent to which a node is influential.

In this section, we propose a new centrality measure for finding the top k influential nodes of the network. We first define a new parameter called the "effective degree" for each node.

Definition 1 *The effective degree of a node i is calculated as follows:*

$$EDeg(i) = outdegree(i) - indegree(i) \quad (6)$$

We also define another parameter, which we call the degree centrality. This parameter measures the local social influence, and its value for each node is calculated based on the sum of that node's effective degree, plus the effective degree of the nodes.

Definition 2 *The degree centrality of node i is calculated as follows:*

$$C_i = EDeg_i + \sum_{j \in \text{jout-neighborsof } i} EDeg_j \quad (7)$$

We prefer to work with smaller values and for all the calculated degree centralities to fall within a positive interval; we therefore subtract the minimum degree centrality that exists in the network from the degree centrality value of each node:

$$C_i = C_i - mi(C_j)(j \in V) \quad (8)$$

where V refers to the set of all nodes in the network.

So far, all we have done has been the calculation of the centrality of nodes based on their immediate neighbours' degrees. However, we now propose a more generic metric that is able to find the centrality based not only on each node's immediate neighbours, but also the structure of the entire graph. In doing so, we define a recursive equation. We refer to the centrality obtained by this recursive equation as the global centrality. This is due to the fact that each node that has neighbours with higher centrality has higher centrality itself.

To calculate this global centrality measure, we first consider the graph $G = (V, E)$, in which $N = |V|$ and $L = |E|$ indicate the number of nodes and edges, respectively. We represent this graph with a $N \times N$ adjacency matrix A .

Each element in the matrix A can either be *oneorzero*; if a_{ij} equals one, this indicates a direct connection between node i and j , whereas a value of zero implies that no connection exists between these two nodes. We normalise the elements of each row by dividing them by the sum of the elements of their corresponding row, in order to transform the adjacency matrix into the form of a stochastic matrix.² We use the stochastic matrix property to prove that our model in Equation 17 converges to the true value.

$$W(i, j) = \frac{A(i, j)}{\sum_k A(i, k)} \quad (9)$$

Obviously, the sum of each row of matrix W is equal to one, and thus the matrix is in the form of a stochastic matrix.

Using this stochastic matrix, we propose a measure for computing the global centrality. Unlike the degree centrality in Eq. 7, the stochastic matrix considers the structure of the entire graph.

$$C^g = \alpha C + (1 - \alpha)WC^g \quad (10)$$

Or in another form

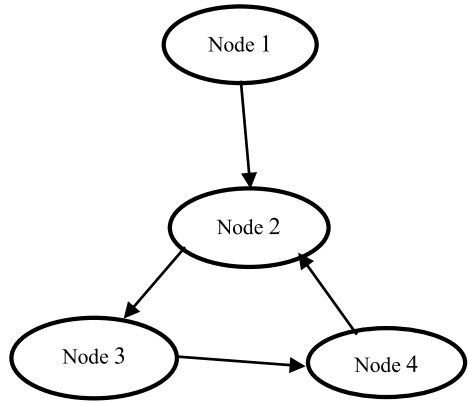
$$C^g = (I - (1 - \alpha)W)^{-1} \alpha C \quad (11)$$

In these equations, C represents the local centrality, or what we have previously defined as the degree centrality in Eq. 7. The parameter α refers to a localisation parameter that optimises the extent to which C and WC^g influence each node's global centrality. α is the value that controls the tradeoff between the global and the local views of the graph. If the value of α is close to one, our computation will be closer to the local centrality, and if it is close to zero, the global centrality will be mainly based on the global view of the whole graph.

We compare our proposed method with existing metrics in order to find the influential nodes. We also introduce an opinion formation model and show that our method can find opinion leaders in the network more effectively and can speed up the process of spreading ideas. We illustrate the performance of our method compared to other methods using a simple example. Consider the graph shown in Fig. 1: node 1 influences the other nodes, but does not

² A matrix in which the values of each of the rows sum to one.

Fig. 1 A sample graph for comparison of centrality measures



follow any other node. For this reason, it is considered to be the best candidate for spreading an idea across the network. We implemented and discussed the centrality measure algorithms for this graph, and the results are presented in Table 1. As can be observed from Table 2, our method identifies node 1 as a leader, while other methods identify node 2 as a leader. The Cg column denotes our proposed algorithm.

3.3 The Degroot Opinion Formation Model

In this section, we present a brief review of the Degroot model [15]. We assume that time is discrete $t = 0, 1, 2, \dots$ and that each agent initially has a predefined opinion x_i concerning some topic. The opinions are values between zero and one, i.e. $x_i \in [0, 1]$. The opinions of all agents at time t are collected in $x(t) \in R$ (we assume n agents). In each period, agents talk to each other, and finally each of the agents updates its own opinion by obtaining a weighted average of the opinions of its neighbours. We then have:

$$x_i(t + 1) = \sum_j T_{ij} x_j(t) \tag{12}$$

where the weight matrix $T = [T_{ij}]_{i,j \in N}$ represents the level of truth that any agent places on the opinion of any other agent. All the elements of T satisfy $t_{ij} \in [0,1]$ for $0 \leq i, j < n$. The elements in each row of the matrix T sum to one, which implies that it is row stochastic. Another variation of this formula can be expressed as:

$$x(t + 1) = Tx(t) = T^{t+1}x(0) \tag{13}$$

Table 2 Comparison between centrality measures in a sample graph

NodeID	Closeness	Betweenness	Eigenvector	PageRank	Cg
Node 1	0.6	0	0.32	0.11	1.0
Node 2	1.0	1.0	1.0	1.0	0.82
Node 3	0.75	0	0.67	0.96	0.89
Node 4	0.75	0	0.67	0.9	0.7

Values are normalised

The weight matrix T can be interpreted as representing the credibility or trust that each agent assigns to the opinions of their neighbors in the network.

An interesting variation of this model was developed by Friedkin and Johnsen [21], Jia et al. [30]. Their model addresses opinion formation under social influence and assumes that an agent i adheres to its initial opinion with a certain probability g_i , and that the agent is socially influenced by the other agents with probability $1 - g_i$. According to the classical model, we have:

$$x_i(t+1) = g_i x_i + (1 - g_i)(a_{i1}(t) + \dots + a_{in}(t)) \quad (14)$$

or in matrix notation,

$$x(t+1) = Gx(0) + (I - G)Ax(t) \quad (15)$$

Here, G is the diagonal matrix where $0 < g_i < 1$ on the diagonal and I is the identity matrix.

3.4 Proposed Opinion Formation Model

In this section, we present a brief overview of the opinion formation model that we intend to use in our benchmarking alongside the Degroot opinion formation model. This model has a different structure from the opinion-leader-finding algorithms, with a different definition of global centrality. In other words, this opinion formation model is one of the novel aspects of this research. Consider a network consisting of n agents (nodes) in which the agents interact with each other. Each agent holds certain opinions on specific topics, and their opinions are updated on a regular basis based on their neighbours' opinions. The issues of how to update opinions and in which situations each agent's opinion is updated are referred to as the opinion formation model in the literature.

Opinion formation models that have been presented so far are either applicable only to undirected graphs [29] or do not take into account the differences between ingoing and outgoing links [8, 17, 54]. In this section, we present a new opinion formation model that takes into account the differences between ingoing and outgoing links.

Consider a graph with n nodes in which each node corresponds to an agent and each neighbourhood relation between two nodes is represented by an edge. For example, i is j 's neighbour if and only if a directed edge (i, j) exists. We define a weight matrix W , and use this matrix to calculate the new opinion of each node. Its elements indicate how much influence each node experiences from its neighbours. The opinion of agent i is represented as x_i and is updated by a linear combination of its neighbours' opinions. It can be calculated as follows:

$$x_i = d_{ii}x_i(0) + \sum_{j=1, j \neq i}^n w_{ij}x_j \quad (16)$$

In matrix form, we have:

$$X = WX + DX. \quad (17)$$

or in another form:

$$X = (I - W)^{-1}DX. \quad (18)$$

As explained earlier, W refers to the influence matrix, in which the elements on the main diagonal are all zero and the other elements represent each neighbour’s influence on a node. Matrix D is a diagonal matrix in which the elements of the main diagonal represent the level of coherence of each agent regarding its own opinion.

The main problem in this model is how to set the values of the influence matrix W . In this research study, we assume that each node has an influence proportional to its degree centrality, which is calculated using the following formula:

$$d_{ii} = \frac{C_i}{\sum_{j \in \text{in-neighbours of } i} C_j + C_i} \tag{19}$$

$$w_{ik} = \frac{C_k}{\sum_{j \in \text{in-neighbours of } i} C_j + C_i} \tag{20}$$

In cases where we have a Eulerian graph,³ which has equal in-degree and out-degree for each node, we set the corresponding C_i zero values to a value of 1/100 to ensure that the matrix $(I - W)^{-1}$ is invertible.

Our opinion formation model is a recursive model, and we build the opinion of each node based on the opinion of other nodes. In this model neighbourhood feature selection selects influential nodes in a network by weighting each node’s opinion influence on its neighbours, considering both ingoing and outgoing links. This method utilizes a weighted matrix W based on degree centrality C_i to measure node influence. In Sect. 2.3.1, we prove that this equation converges to a unique answer. In Sect. 2.3.2, we presented a method of finding this unique answer.

3.4.1 Solving the Opinion Formation Model

Equation 17 has the form of a system of linear equations, and we can solve this by various direct and iterative methods. Cramer, Gaussian elimination, Gaussian Jordan elimination, and inverse matrix computation are examples of direct methods that are not suitable for problems with large numbers of variables due to their high computational complexities. For example, the inverse matrix method has a time complexity of $O(n^3)$, and the Cramer method has a complexity of $O(n!)$, making these methods impractical for problems with more than three variables. We need to solve the following equation:

$$X = WX + DX. \tag{21}$$

In addition, in matrix form, we have

$$\begin{aligned} [x_1 x_2 \dots x_n] &= \left[0 \frac{w_{12}}{\sum_k w_{1k}} \frac{w_{13}}{\sum_k w_{1k}} \dots \frac{w_{1n}}{\sum_k w_{1k}} \frac{w_{21}}{\sum_k w_{2k}} 0 \frac{w_{23}}{\sum_k w_{2k}} \dots \frac{w_{2n}}{\sum_k w_{2k}} \dots \dots \right] \\ &\quad \left[\frac{w_{n1}}{\sum_k w_{nk}} \frac{w_{n2}}{\sum_k w_{nk}} \frac{w_{n3}}{\sum_k w_{nk}} \dots 0 \right] [x_1 x_2 \dots x_n] \\ &\quad + \left[\frac{w_{11}}{\sum_k w_{1k}} 0 0 \dots 0 0 \frac{w_{22}}{\sum_k w_{2k}} 0 \dots 0 \dots \dots 0 0 \dots \frac{w_{nn}}{\sum_k w_{nk}} \right] \\ &\quad [x_1^{(e)} x_2^{(e)} \dots x_n^{(e)}] \end{aligned} \tag{22}$$

³ Directed graphs in which every node has the same in-degree and out-degree.

Our research study is mainly performed on large graphs such as the Facebook and Twitter networks with high numbers of variables; we therefore need methods with a linear time complexity. In view of this difficulty, we can use iterative methods such as Jacobi and Gauss-Siedel methods, since these methods converge with linear time complexities. Here, we use the Jacobi iterative method. As a result, the iterative solution for each step of the algorithm is formulated as follows:

$$x_i^{(t)} = d_{ii}x_i^{(s)} + \sum_{j=1, j \neq i}^n w_{ij}x_j^{(t-1)} \quad (23)$$

3.5 Finding Opinion Leaders: An Algorithmic Approach

The global centrality defined in the previous section forms the foundation of our algorithm. We use the algorithm schema proposed by Salehi-Abari and Boutilier [50] and apply our centrality measure to find the opinion leaders. We need to find the nodes with most influence, which is similar to the global centrality. The influence rate for each node is computed using a recursive equation based on its neighbours' influence rates. In this definition, nodes that are connected to more influential nodes are more influential themselves.

Based on the definition of global centrality presented in Eq. 10, we define the influence computation formula as follows:

Definition 3 *The influence of node i is calculated as follows:*

$$Influ_i = \alpha C_i + (1 - \alpha) \sum_{j \neq i} w_{ij} Influ_j \quad (24)$$

or in matrix form:

$$Influ = \alpha C + (1 - \alpha) W Influ \quad (25)$$

or in another form

$$Influ = (I - (1 - \alpha)W)^{-1} \alpha C \quad (26)$$

where $Influ$ is the influence vector and C is the degree centrality defined above in Eq. 7. The iterative form of this equation in the $t + 1$ th step is

$$Influ_i^{(t+1)} = \alpha C_i + (1 - \alpha) \sum_{j \neq i} w_{ij} Influ_j^{(t)} \quad (27)$$

Moreover, in matrix form, we have

$$Influ^{(t+1)} = \alpha C + (1 - \alpha) W Influ^{(t)} \quad (28)$$

We set the initial influence value equal to the effective degree ($Influ^{(0)} = EDeg$), as introduced in Def. (6).

Corollary 1 *Equation 27 has a unique solution and the iterative approach converges to a correct solution.*

Proof In order to provide a proof for this, we need to prove that $1 - \alpha w$ is invertible. As explained above, W is a row stochastic matrix, so the sum of the elements in each row sums to one and $\alpha < 1$. Therefore, $1 - \alpha w < 1$ and is invertible.

In order to design an algorithm based on this approach, we define an error interval. We use this error interval to eliminate some of the nodes at each step of the algorithm that do not influence the final result. To proceed further in computing this interval, we need to use the definitions of norms, as follows.

Definition 4 A scalar norm is a type of norm that is defined for a vector or matrix and has the following properties:

1. For each $x \in R^n \times n, \alpha \in R$ we have $\|\alpha x\| = |\alpha| \|x\|$
2. $\|XY\|_p \leq \|X\|_p \|Y\|_p$
3. $\|X^k\|_p \leq \|X\|_p^k$

Definition 5 An infinite norm is defined as follows:

$$\|A\|_\infty = \max_i \left(\sum_{j=1}^n |a_{ij}| \right) \tag{29}$$

Theorem 1 The error value in the t th step of the proposed iterative method is bounded by the following interval:

$$\|Influ - Influ^t\|_\infty \leq (1 - \alpha)^t \|Influ - Influ^\circ\|_\infty \tag{30}$$

Influ is the correct value of the influence, which is calculated by the proposed iterative method, and $Influ^\circ$ represents the initial value of the influence.

Proof *Influ* is the correct value of the influence after convergence of the proposed iterative approach. We therefore have:

$$Influ = (1 - \alpha)WInflu \tag{31}$$

We also have

$$\|Influ - Influ^t\|_\infty = \|(1 - \alpha)WInflu - (1 - \alpha)WInflu^{(t-1)}\|_\infty \tag{32}$$

which is equal to:

$$= \|((1 - \alpha)W)^t (Influ - Influ^\circ)\|_\infty \tag{33}$$

From Property 2 of Definition 4, we have:

$$\leq (1 - \alpha)^t \|W^t\|_\infty \|Influ - Influ^\circ\|_\infty \tag{34}$$

In addition, from Property 3 of Definition 4, we can write:

$$\leq (1 - \alpha)^t W_\infty^t \|Influ - Influ^\circ\|_\infty \tag{35}$$

By using the property of the infinite norm:

$$= (1 - \alpha)^t \left(\sum_{j=1}^n |w_{ij}| \right)^t \|Influ - Influ^\circ\|_\infty \tag{36}$$

which yields the following result:

$$= (1 - \alpha)^t \|Influ - Influ^\circ\|_\infty \tag{37}$$

Using this theorem, we can find the most influential nodes at each step.

Theorem 2 For each node in the t th iteration, the maximum value of the error is bounded by the following interval.

$$\left| Influ_i - Influ_i^{(t)} \right| \leq (1 - \alpha)^t ((Influ^\circ) - \min(Influ^\circ)) \tag{38}$$

Proof From theorem 1, we have:

$$\|Influ - Influ^{(t)}\|_\infty \leq (1 - \alpha)^t \|Influ - C\|_\infty \tag{39}$$

From the properties of the infinite norm:

$$\left| Influ_i - Influ_i^{(t)} \right| \leq (1 - \alpha)^t |Influ - C| \tag{40}$$

$$\left| Influ_i - Influ_i^{(t)} \right| \leq (1 - \alpha)^t \left| Influ - Influ^{(c)} \right| \tag{41}$$

Matrix W in Eq. 27 is normalised. A node’s influence in each iteration is therefore a linear combination of the normal coefficients of the influences of all nodes, and will never violate its maximum. Thus, we always have $\max(influ) \leq (influ^\circ)$. In addition, as a result of the equality of the initial influence and the degree centrality vector ($influ^\circ = C$), we have $\max(influ) \leq \max(C)$. We can therefore conclude:

$$\left| (Influ - Influ^\circ) \right| \leq (influ^\circ) - \min(influ^\circ) \tag{42}$$

$$\left| influ_i - influ_i^{(t)} \right| \leq (1 - \alpha)^t ((C) - (C)) \tag{43}$$

Using the above theorem, we can conclude the following proposition:

Proposition 1 If $influ_i^{(t)} - influ_j^{(t)} > 2(1 - \alpha)^t ((C) - \min(C))$ then $influ_i > influ_j$

Proof $influ_i^t - influ_j^t = influ_i^t + influ_i - influ_i - influ_j^t + influ_j - influ_j$

$$= |influ_i - influ_i^t| + influ_i + \left| influ_j - influ_j^t \right| - influ_j \tag{44}$$

From Theorem 5, we have:

$$influ_i^t - influ_j^t \leq 2(1 - \alpha)^t ((C) - (C)) + influ_i - influ_j \tag{45}$$

According to the proposition itself, we also know that:

$$influ_i^{(t)} - influ_j^{(t)} > 2(1 - \alpha)^t ((C) - \min(C)) \tag{46}$$

With respect to this inequality, we know:

$$\text{influ}_i^t - \text{influ}_j^t \leq 2(1 - \alpha)^t((C) - (C)) + \text{influ}_i - \text{influ}_j \quad (47)$$

To make both of these inequalities true, we need:

$$\text{influ}_i - \text{influ}_j < 0 \quad (48)$$

We therefore have:

$$\text{influ}_i > \text{influ}_j. \quad (49)$$

Using this error interval, we can design a time-efficient algorithm for finding the top k influential nodes in the network. In this algorithm, the initial influence of each node is set to the degree centrality, as discussed above. At each iteration, the value of the influence of each node is updated based on Eq. 27, and the nodes that do not satisfy the conditions of Proposition (1) are eliminated.

The process of eliminating these nodes is as follows: in each iteration of the algorithm, we find the k th influential nodes which have a different in influence from the detected node that is greater than the threshold in Proposition (1); these are considered not to be influential nodes. The exclusion of less influential nodes in the current iteration is a heuristic used to reduce the computational complexity of the algorithm. We can be sure that these nodes are not in the set of the top k most influential nodes in the current iteration, and there is therefore a good chance that they will not be in the same set in the future. This hypothesis is supported with good experimental results. We have also built our algorithm on the assumption that nodes with a smaller number of influential neighbours are less influential themselves. Hence, there is a relatively small probability that the exclusion of less effective neighbours at each iteration will lead to a profound difference in the final results, and thus they are extracted from the set of nodes.

4 Experimental Results

In this section, we demonstrate the practical performance of our approach on several datasets collected from real-world social networks. Our method is evaluated based on both the Degroot opinion formation model and our opinion formation model, as discussed above. We use three variations of the Degroot model, as explained below. The reason for using the Degroot model in addition to our own opinion formation model is to prevent the assumption of a ‘chicken and egg’ problem that might arise due to the similarities between our model and an opinion leader-finding algorithm. In other words, we show that our model is not the only one to give good performance using the algorithm’s outputs.⁴ However, our own opinion formation model shows good performance for our algorithm, and in a real-world situation in which the ‘chicken and egg’ problem is not a concern, our model and opinion-finding algorithm can be used simultaneously. The statistics of all datasets are summarised in Table 2, in which the parameters N and M denote the number of nodes and edges in the network, k is the average degree of the nodes in the network, and k_{max} is the maximum degree of a node that exists in the network. The parameter C represents the average of the clustering

⁴ The datasets can be downloaded from: <http://konnect.uni-koblenz.de/networks/>.

Table 3 Properties of test networks

Network	N	M	$\langle k \rangle$	k_{max}	C	LCC
Advogato	6541	51,127	15.633	943	9.22	5042
Epinions	75,879	508,837	13.412	3079	6.57	75,877
Twitter	465,017	834,797	3.5904	678	0.0613	465,017
Pokec	1,632,803	30,622,564	37.509	20,518	4.68	1,632,803

coefficient of the nodes in the graph, and LCC represents the size of the largest connected component in each graph.

We evaluated our approach on four datasets: Advogato, Epinions, Pokec and Twitter. The Advogato network [42] is an online community platform launched in 1999, catering specifically to developers of free software. In this network, nodes represent individual users of Advogato, and directed edges signify trust relationships between these users, reflecting the degree of trust each user places in others within the community. The Epinions network [49] is derived from an online social network where nodes represent users of the platform. Directed edges in this network denote trust relationships established between users, indicating which users trust others' opinions or recommendations on products and services within the platform. The Pokec network [55] originates from the Slovak social network Pokec, where nodes correspond to individual users of the platform. Directed edges in this friendship network represent mutual friendships between users, highlighting social connections and interactions among members of the Pokec community. Twitter [19] is a directed network where nodes represent users of the platform, and each directed edge indicates that the user on the left-hand side follows the user on the right-hand side. This network structure captures the follower relationships among users, reflecting the influence and connectivity dynamics within the Twitter platform.

Firstly, to test the influence of the detected influential nodes in the opinion formation model, a random value in the range $[0, 1]$ is assigned to each node to represent its opinion on certain subjects. If this value is close to zero, this indicates that the node has a negative view on that subject; conversely, if this value is close to one, this indicates that this node fully supports this subject. In all our experiments, we consider 10% of the nodes to be influential, and thus their opinions are set to one. We update each node's opinion based on the opinion formation model used in the experiment until the results converge to certain values and no further changes are seen. The final opinion in the network, which is the average of all nodes' opinions, is used as a benchmark to evaluate the influence of influential nodes.

We conduct these tests for three different scenarios: (i) all initial values are set to 0; (ii) all initial values are set to 0.5; and (iii) the initial values are random, i.e. the values are set to a random number between zero and one. The value for parameter α is set to 0.8, since the algorithm works best for this value, as observed in Sect. 3.4.

4.1 Test Results for the Proposed Opinion Formation Model

In this section, the efficiency of the proposed method using the opinion formation model is compared to other commonly used methods, namely betweenness centrality, closeness centrality, eigenvector centrality and PageRank centrality. We assign initial values as explained at the beginning of Sect. 4. We run the test for our proposed opinion formation model and compare the values of the opinions for each method at each iteration. Thus, the results presented here are the averages of 20 runs of the simulation.

In these experiments, the opinions of the users were numbers in the range 0–1; if the value is close to one, then it means that a particular user supports the idea completely; its closeness to zero indicates the opposite. The y-axis in the figures denotes the average of the opinions of all nodes, and the x-axis indicates the time step in the test.

4.2 Test Results for the Degroot Opinion Formation Model

As in the previous section, we test our proposed method and compare the results to other methods; however, this time our tests are conducted for the Degroot model. All other evaluation criteria are the same as in Sect. 4.1.

4.3 Test Results for Three Different Variations of Centrality for the Degroot Opinion Formation Model

In this section, we compare different choices for C_i (here we use C_i instead of T_{ij} , which was used in the original Degroot model) in Eq. 12 for the Degroot opinion formation model. The Degroot model is used for all of these tests. The three variations of C_i are as follows: (i) using global centrality, as in Sects. 4.1 and 4.2; (ii) including nodes that are two hops away in computing C_i with a discount factor of 0.5; and (iii) using degree imbalances⁵ without adding the neighbours' degree imbalances (Table 3).

⁵ Absolute values of the in-degree minus the out-degree of nodes.

Algorithm 1 Mining Top k Most influential Nodes

```

1: input: network  $G = (V, E)$ , size of result  $k$ 
2: output: Top- $k$  influential nodes;
3: For each vertex  $i$  do
4:    $influ_i^0 \leftarrow outdegree(i) - indegree(i)$ 
5: end for
6:  $t \leftarrow 0$ 
7: while  $size(V) > k$  do
8:    $t \leftarrow t + 1$ 
9:   for  $i \leftarrow 1$  to  $size(V)$  do
10:     $influ_i^{t+1} \leftarrow \alpha influ_i^0 + (1 - \alpha) \sum_{j \text{ is in-neighbors of } i} w_{ij} influ_j^t$ 
11:   end for
12:    $kth\ max^t \leftarrow kth\ largest\ element\ in\ influ_{v \in V}^t$ 
13:    $influ_{max}^t \leftarrow \max_{v \in V} (influ_v^t)$ 
14:    $influ_{min}^t \leftarrow \min_{v \in V} (influ_v^t)$ 
15:   for each vertex  $v$  in  $V$  do
16:     if  $kth\ max^t - influ_v^{(t)} > 2(1 - \alpha)^t (influ_{max}^t - influ_{min}^t)$  then
17:        $V \leftarrow V - v$ 
18:     end if
19:   end for
20: end while
21: return  $V$ 

```

4.3.1 Evaluation of Results

As we can observe from Fig. 2 and from all the datasets except the Epinions network, the values of the opinions increase and finally converge to a particular value. In the Epinions dataset, the values of the opinions decrease but also converge as in the other datasets. We can see the same behaviour in Figs. 3 and 4 in all datasets (except for the Epinions network in Fig. 3). In all cases, the proposed method was superior to the other methods of finding

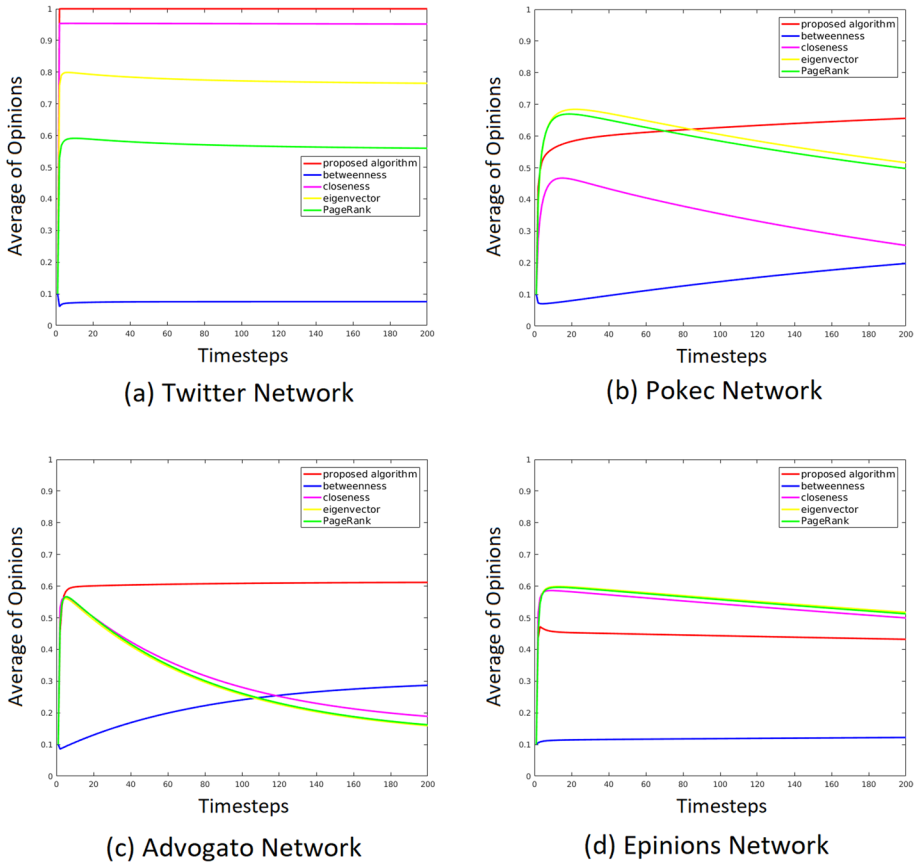


Fig. 2 Comparison of methods for finding opinion leaders in the Degroot opinion formation model with global centrality and initial values of zero

opinion leaders, except for the case of the Epinions network in Fig. 2 and the Pokec network in Fig. 4. As can be seen from Fig. 5, our method outperforms the other methods of finding opinion leaders since it considers both influence and conformity simultaneously. In the Pokec network, the proposed method gives the same performance as the betweenness centrality and is better than the other choices.

Another interesting observation is that there is no significant difference between the curves except for Pokec network in Figs. 2, 3 and 4 and the Epinions network in Fig. 2. Both the results for the proposed opinion formation model and for the Degroot model converge immediately to almost the same curves, regardless of the method used. This is due to the fact that if the initial opinion leaders are chosen correctly, then there is no significant difference of what opinion formation model they have and this will converge to the same results. The consistent convergence of opinion values across Figs. 2, 3, and 4 reveals important insights for abnormal behaviour analysis using truth discovery methods. This can be used to leverage weighted matrices for representing trust and influence, where deviations from expected convergence patterns may indicate anomalies or influential nodes exhibiting atypical behaviours. Such an approach not only improves outlier detection but

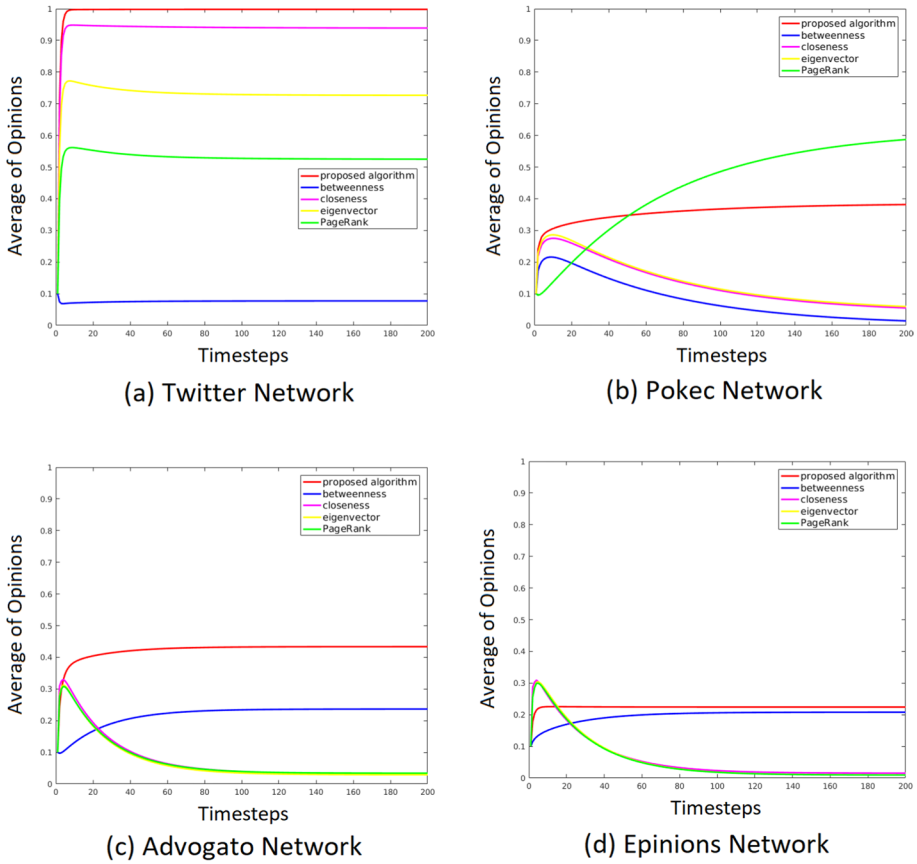
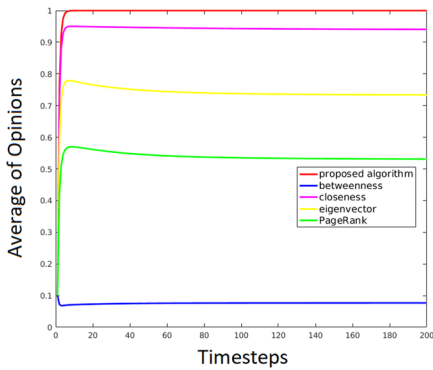


Fig. 3 Comparison of methods for finding opinion leaders in the Degroot opinion formation model with two-hop centrality and initial values of zero

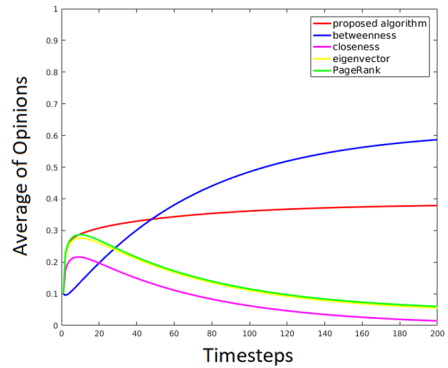
also deepens our understanding of how information propagates and exerts influence within intricate social networks.

4.4 Fault Tolerance Against Noisy Data

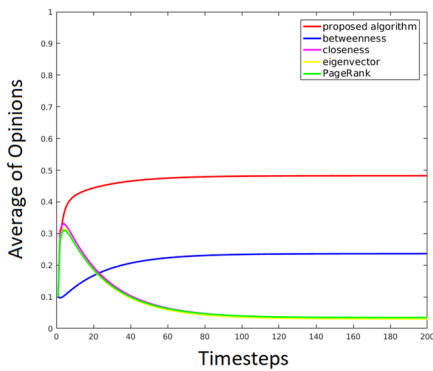
Fault tolerance is of paramount importance, especially in situations where some of the links in the input data do not exist in the real world and our data contain errors. Social networking data may be associated with uncertainty, especially in cases where we determine relationships between agents. For example, in a protein network obtained from a biological experiment, there is always the possibility of error when modelling the links of the network. In order to achieve the task of finding the top k influential nodes in these networks, we require algorithms with great tolerance against such errors. In this section, we examine the fault tolerance of our proposed method against the PageRank algorithm. In order to create a graph containing errors, we add some edges to our original graph. The addition of wrong connections to the graph makes the graph ‘noisy’. In the next



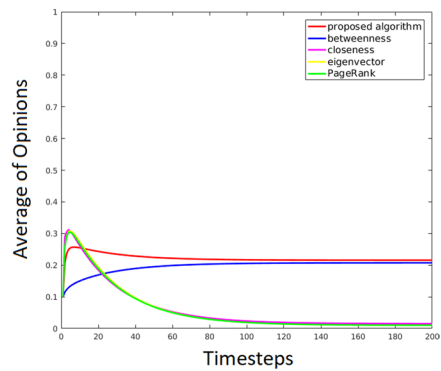
(a) Twitter Network



(b) Pokec Network



(c) Advogato Network



(d) Epinions Network

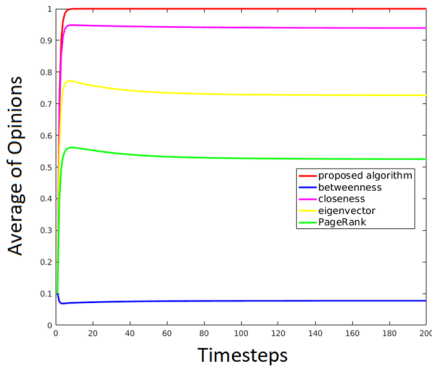
Fig. 4 Comparison of methods for finding opinion leaders in the Degroot opinion formation model with degree imbalance as centrality with initial values of zero

step, we compute the centrality for this noisy graph and measure the error between the noisy and normal graphs using the following equation:

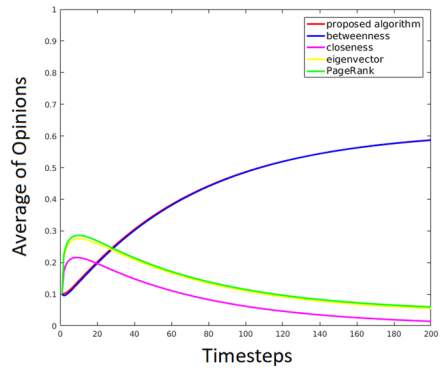
$$I_s = \sum_{i=1}^N |S_i^t - S_i| \quad (50)$$

S_i^t and S_i correspond to the centrality score of node i in the noisy and original graphs, respectively. In Fig. 6, the term ‘spurious edges’ refers to the number of edges that do not actually exist in the graph. As previously stated, these are errors in the network. As shown in the same figure, our centrality is more robust to noisy data than the PageRank centrality, as indicated by its lower value of I_s .

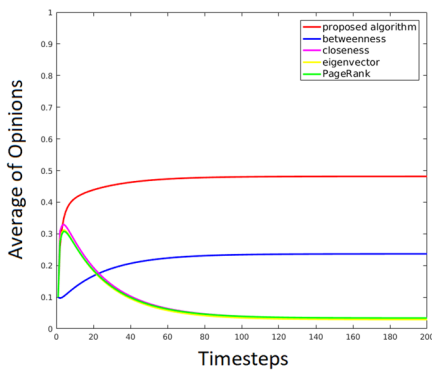
Figure 7 shows the final opinions in the network that are reached by multiple proportions of leaders in the Pokec network. This curve has an ascending shape, and by increasing the number of leaders, the average of the final opinion of the system increases. The results for the other networks tested are similar. In Fig. 8, we can observe



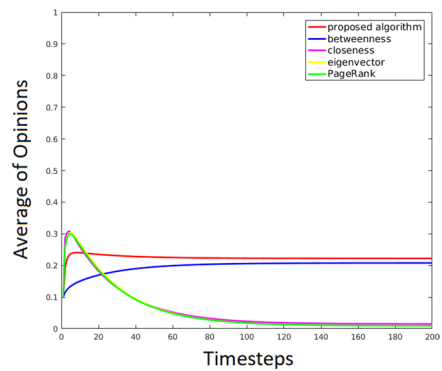
(a) Twitter Network



(b) Pokec Network



(c) Advogato Network



(d) Epinions Network

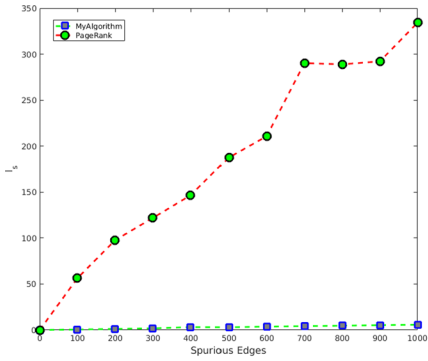
Fig. 5 Comparison of methods for finding opinion leaders in the proposed opinion formation model with initial values of zero

the effects of different values of α on the average of the opinions of the nodes in the Pokec network. A larger value of α makes leaders more local, while a smaller value makes them more global. The results for the other networks tested are similar.

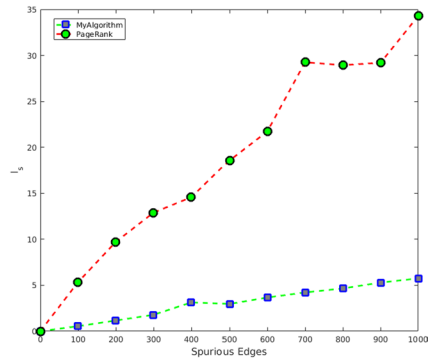
5 Conclusion and Future Works

In this research paper, we propose a new centrality measure to find opinion leaders and an algorithm to find these leaders. Furthermore, we introduce a new opinion formation model to calculate the opinions of the network nodes.

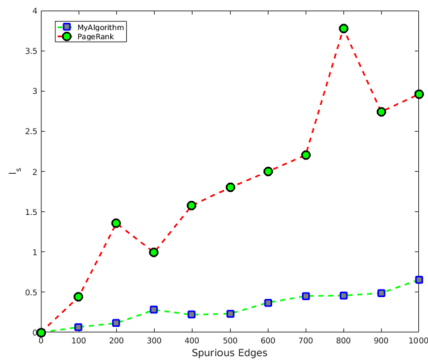
The main advantage of our method is that it considers both the input and output links, while previous methods have mostly focused on the output links and have ignored the role of the input links. We test our proposed opinion-leader-finding algorithm with our proposed opinion formation model and the Degroot opinion formation model. Our method outperforms classic centrality measures, and also has the advantage of fault tolerance.



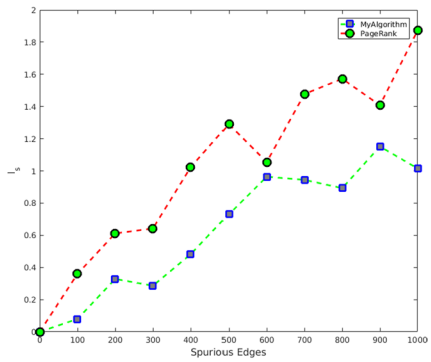
(a) Twitter Network



(b) Pokec Network



(c) Advogato Network



(d) Epinions Network

Fig. 6 Comparison of fault tolerance in the proposed method and the PageRank method

Fig. 7 Effect of the proportion of leader nodes on the average of the final opinions of nodes in the Pokec network

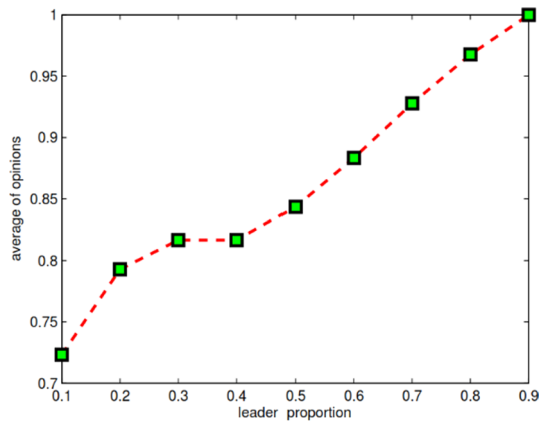
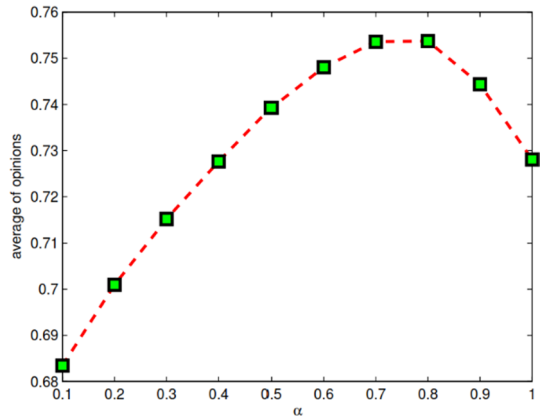


Fig. 8 Effect of different values of α on the average of the final opinions of nodes in the Pokec network



Our proposed algorithm is specifically designed to identify a small subset of nodes within a social network that can significantly influence the opinions of other nodes. By leveraging our novel centrality measure and opinion formation model, we can pinpoint these key nodes with greater accuracy and efficiency than traditional methods. This approach not only enhances the precision of identifying opinion leaders but also underscores the importance of considering the bidirectional nature of interactions within the network. Our method is evaluated against the Degroot opinion formation model and our own model, demonstrating superior performance in large real-world social networks such as Facebook, Twitter, Advogato, and Epinions. We tested the influence of detected nodes by assigning initial opinion values and updating them iteratively until convergence, showing that our model leads to more accurate identification of influential nodes compared to traditional centrality measures like betweenness, closeness, eigenvector, and PageRank centralities.

Directions for future research may include the implementation of our method in a distributed framework (e.g. Hadoop, using the MapReduce programming model). Another interesting area of research would be the application of these methods to graphs for which we do not have full knowledge about the network's structure, or to graphs in which the topology of the graph changes constantly, thus giving the network a dynamic structure.

Author Contributions First author contributed to the implementation and second author contributed to writing and preparing the technical soundness for publication. Third author was the supervisor and revised the manuscript and helped in idea generation and supervised both first and second authors.

Funding The authors received no funding for this work.

Data Availability All the data for this work are publicly available as cited in the text Sect. 4.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Ali, A., Hulipalled, V. R., & Patil, S. S. (2020). Centrality measure analysis on protein interaction networks. In *Proceedings of 2020 IEEE international conference on technology, engineering, management for societal impact using marketing, entrepreneurship and talent, TEMSMET 2020*. <https://doi.org/10.1109/TEMSMET51618.2020.9557447>
2. Alshahrani, M., Fuxi, Z., Sameh, A., Mekouar, S., & Huang, S. (2020). Efficient algorithms based on centrality measures for identification of top-K influential users in social networks. *Information Sciences*, 527, 88–107. <https://doi.org/10.1016/j.ins.2020.03.060>
3. Bao, Z.-K., Ma, C., Xiang, B.-B., & Zhang, H.-F. (2017). Identification of influential nodes in complex networks: Method from spreading probability viewpoint. *Physica A: Statistical Mechanics and its Applications*, 468, 391–397.
4. Bloch, F., Jackson, M. O., & Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, 61(2), 413–453. <https://doi.org/10.1007/s00355-023-01456-4>
5. Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5), 1170–1182.
6. Bordignon, V., Matta, V., & Sayed, A. H. (2021). Adaptive social learning. *IEEE Transactions on Information Theory*, 67(9), 6053–6081. <https://doi.org/10.1109/TIT.2021.3094633>
7. Borgs, C., Brautbar, M., Chayes, J., & Lucier, B. (2014). Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on discrete algorithms* (pp. 946–957).
8. Buechel, B., Hellmann, T., & Klößner, S. (2015). Opinion dynamics and wisdom under conformity. *Journal of Economic Dynamics and Control*, 52, 240–257.
9. Cai, T., Li, J., Mian, A., Li, R. H., Sellis, T., & Yu, J. X. (2022). Target-aware holistic influence maximization in spatial social networks. *IEEE Transactions on Knowledge and Data Engineering*, 34(4), 1993–2007. <https://doi.org/10.1109/TKDE.2020.3003047>
10. Candeloro, L., Savini, L., & Conte, A. (2016). A new weighted degree centrality measure: The application in an animal disease epidemic. *PLoS ONE*, 11(11), e0165781.
11. Chen, D.-B., Gao, H., Lü, L., & Zhou, T. (2013). Identifying influential nodes in large-scale directed networks: The role of clustering. *PLoS ONE*, 8(10), e77455.
12. 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. <https://doi.org/10.1016/j.techfore.2011.06.003>
13. Curado, M., Tortosa, L., & Vicent, J. F. (2023). A novel measure to identify influential nodes: Return random walk gravity centrality. *Information Sciences*, 628, 177–195. <https://doi.org/10.1016/j.ins.2023.01.097>
14. Das, K., Samanta, S., & Pal, M. (2018). Study on centrality measures in social networks: A survey. *Social Network Analysis and Mining*, 8(1), 13.
15. Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118.
16. Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, 58(10), 1260–1277.
17. Eom, Y.-H., & Shepelyansky, D. L. (2015). Opinion formation driven by PageRank node influence on directed networks. *Physica A: Statistical Mechanics and Its Applications*, 436, 707–715.
18. Fei, L., Mo, H., & Deng, Y. (2017). A new method to identify influential nodes based on combining of existing centrality measures. *Modern Physics Letters B*, 31, 1750243.
19. Fensel, D., Sycara, K., & Mylopoulos, J. (2003). *LNCS 2870—The Semantic Web—ISWC 2003*.
20. Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40, 35–41.
21. Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3–4), 193–206.
22. Ghafouri, S., Khasteh, S. H., & Azarkasb, S. O. (2024). Influence maximization (IM) in complex networks with limited visibility using statistical methods. *Journal of Supercomputing*, 80(5), 6809–6854. <https://doi.org/10.1007/s11227-023-05695-1>
23. Guo, J., Zhang, P., Zhou, C., Cao, Y., & Guo, L. (2013). Personalized influence maximization on social networks. In *Proceedings of the 22nd ACM international conference on information & knowledge management* (pp. 199–208).
24. Herrmann, D. A. (2022). Prediction with expert advice applied to the problem of prediction with expert advice. *Synthese*, 200(4), 315. <https://doi.org/10.1007/s11229-022-03809-5>

25. Hou, L. (2022). Network versus content: The effectiveness in identifying opinion leaders in an online social network with empirical evaluation. *Physica A: Statistical Mechanics and Its Applications*, 592, 126879. <https://doi.org/10.1016/j.physa.2022.126879>
26. Jain, L., Katarya, R., & Sachdeva, S. (2020). Opinion leader detection using whale optimization algorithm in online social network. *Expert Systems with Applications*, 142, 113016. <https://doi.org/10.1016/j.eswa.2019.113016>
27. Jain, L., Katarya, R., & Sachdeva, S. (2020). Recognition of opinion leaders coalitions in online social network using game theory. *Knowledge-Based Systems*, 203, 106158. <https://doi.org/10.1016/j.knsys.2020.106158>
28. Jain, L., Katarya, R., & Sachdeva, S. (2023). Opinion leaders for information diffusion using graph neural network in online social networks. *ACM Transactions on the Web*, 17(2), 1–17. <https://doi.org/10.1145/3580516>
29. Jalili, M. (2013). Social power and opinion formation in complex networks. *Physica A: Statistical Mechanics and Its Applications*, 392(4), 959–966.
30. Jia, P., MirTabatabaei, A., Friedkin, N. E., & Bullo, F. (2015). Opinion dynamics and the evolution of social power in influence networks. *SIAM Review*, 57(3), 367–397.
31. Joyce, K. E., Laurienti, P. J., Burdette, J. H., & Hayasaka, S. (2010). A new measure of centrality for brain networks. *PLoS ONE*, 5(8), e12200. <https://doi.org/10.1371/journal.pone.0012200>
32. Kayaalp, M., Bordignon, V., & Sayed, A. H. (2024). Social opinion formation and decision making under communication trends. *IEEE Transactions on Signal Processing*, 72, 506–520. <https://doi.org/10.1109/TSP.2023.3347918>
33. Kazemzadeh, F., AsgharSafaei, A., Mirzarezaee, M., Afsharian, S., & Kosarirad, H. (2023). Determination of influential nodes based on the Communities' structure to maximize influence in social networks. *Neurocomputing*, 534, 18–28. <https://doi.org/10.1016/j.neucom.2023.02.059>
34. Kempe, D., Kleinberg, J. M., & Tardos, É. (2015). Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4), 105–147.
35. Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In: *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 137–146).
36. Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11), 888–893. <https://doi.org/10.1038/nphys1746>
37. Kozitsin, I. V. (2022). A general framework to link theory and empirics in opinion formation models. *Scientific Reports*, 12(1), 5543. <https://doi.org/10.1038/s41598-022-09468-3>
38. Li, J., Cai, T., Deng, K., Wang, X., Sellis, T., & Xia, F. (2020). Community-diversified influence maximization in social networks. *Information Systems*, 92, 101522. <https://doi.org/10.1016/j.is.2020.101522>
39. Li, Q., Zhou, T., Lü, L., & Chen, D. (2014). Identifying influential spreaders by weighted Leader-Rank. *Physica A: Statistical Mechanics and Its Applications*, 404, 47–55.
40. Li, Z., Tang, J., Zhao, C., & Gao, F. (2023). Improved centrality measure based on the adapted PageRank algorithm for urban transportation multiplex networks. *Chaos, Solitons and Fractals*, 167, 112998. <https://doi.org/10.1016/j.chaos.2022.112998>
41. Loeper, A., Steiner, J., & Stewart, C. (2014). Influential opinion leaders. *The Economic Journal*, 124(581), 1147–1167.
42. Massa, P., Salvetti, M., & Tomasoni, D. (2009). Bowling alone and trust decline in social network sites. In *8th IEEE international symposium on dependable, autonomic and secure computing, DASC 2009* (pp. 658–663). <https://doi.org/10.1109/DASC.2009.130>
43. Mohseni, A., & Williams, C. R. (2019). Truth and conformity on networks. *Erkenntnis*. <https://doi.org/10.1007/s10670-019-00167-6>
44. Molinero, X., & Riquelme, F. (2021). Influence decision models: From cooperative game theory to social network analysis. *Computer Science Review*, 39, 100343. <https://doi.org/10.1016/j.cosrev.2020.100343>
45. Nguyen, V. X., Xiao, G., Xu, X. J., Wu, Q., & Xia, C. Y. (2020). Dynamics of opinion formation under majority rules on complex social networks. *Scientific Reports*, 10(1), 456. <https://doi.org/10.1038/s41598-019-57086-3>
46. Page, L., & Brin, S. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks*, 30(1–7), 107–117. [https://doi.org/10.1016/s0169-7552\(98\)00110-x](https://doi.org/10.1016/s0169-7552(98)00110-x)
47. Qian, L., Ge, X., Feng, Z., Wang, S., Yuan, J., Pan, Y., Shi, H., Xu, J., & Sun, Y. (2023). Brain network reorganization during visual search task revealed by a network analysis of fixation-related

- potential. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, 1219–1229. <https://doi.org/10.1109/TNSRE.2023.3242771>
48. Rashidi, R., Boroujeni, F. Z., Soltanaghaei, M., & Farhadi, H. (2024). Prediction of influential nodes in social networks based on local communities and users' reaction information. *Scientific Reports*, 14(1), 15815. <https://doi.org/10.1038/s41598-024-66277-6>
 49. Sadagopan, S., Bertino, E., Kumar, R., & Association for Computing Machinery. (2011). In *Proceedings of the 20th international conference on World Wide Web: WWW 11 20th International World Wide Web Conference, Hyderabad, India, March 28–April 01, 2011*. Association for Computing Machinery.
 50. Salehi-Abari, A., & Boutilier, C. (2014). Empathetic social choice on social networks. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems* (pp. 693–700).
 51. Shumovskaia, V., Kayaalp, M., Cemri, M., & Sayed, A. H. (2023). Discovering influencers in opinion formation over social graphs. *IEEE Open Journal of Signal Processing*, 4, 188–207. <https://doi.org/10.1109/OJSP.2023.3261132>
 52. Simko, G. I., & Csermely, P. (2013). Nodes having a major influence to break cooperation define a novel centrality measure: Game centrality. *PLoS ONE*, 8(6), e67159.
 53. Sun, M. W., Moretti, S., Paskov, K. M., Stockham, N. T., Varma, M., Chrisman, B. S., Washington, P. Y., Jung, J. Y., & Wall, D. P. (2020). Game theoretic centrality: A novel approach to prioritize disease candidate genes by combining biological networks with the Shapley value. *BMC Bioinformatics*, 21(1), 356. <https://doi.org/10.1186/s12859-020-03693-1>
 54. Szalay, K. Z., & Csermely, P. (2013). Perturbation centrality and turbine: A novel centrality measure obtained using a versatile network dynamics tool. *PLoS ONE*, 8(10), e78059.
 55. Takac, L., & Zabojsky, M. (2012). *Data analysis in public social networks*. <http://www.pokec.sk>
 56. Tang, Y., Xiao, X., & Shi, Y. (2014). Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on management of data* (pp. 75–86).
 57. Wan, Z., Mahajan, Y., Kang, B. W., Moore, T. J., & Cho, J. H. (2021). A survey on centrality metrics and their network resilience analysis. *IEEE Access*, 9, 104773–104819. <https://doi.org/10.1109/ACCESS.2021.3094196>
 58. Wang, S., Du, Y., & Deng, Y. (2017). A new measure of identifying influential nodes: Efficiency centrality. *Communications in Nonlinear Science and Numerical Simulation*, 47, 151–163.
 59. Wang, W., Nie, Y., Li, W., Lin, T., Shang, M. S., Su, S., Tang, Y., Zhang, Y. C., & Sun, G. Q. (2024). Epidemic spreading on higher-order networks. *Physics Reports*, 1056, 1–70. <https://doi.org/10.1016/j.physrep.2024.01.003>
 60. Wang, Y., Li, H., Zhang, L., Zhao, L., & Li, W. (2022). Identifying influential nodes in social networks: Centripetal centrality and seed exclusion approach. *Chaos, Solitons and Fractals*, 162, 112513. <https://doi.org/10.1016/j.chaos.2022.112513>
 61. Yanchenko, E., Murata, T., & Holme, P. (2024). Influence maximization on temporal networks: A review. *Applied Network Science*, 9(1), 16. <https://doi.org/10.1007/s41109-024-00625-3>
 62. Zino, L., & Cao, M. (2021). Analysis, prediction, and control of epidemics: A survey from scalar to dynamic network models. *IEEE Circuits and Systems Magazine*, 21(4), 4–23. <https://doi.org/10.1109/MCAS.2021.3118100>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Zahra Ghorbani is an M.Sc student at K. N. Toosi, University of Technology, Tehran, Iran, and her research interests are complex networks and social media analysis.



Saeid Ghafouri received the B.Sc. degree in computer software engineering from Shahid Chamran University of Ahvaz, Ahvaz, Iran. He received his M.Sc. in the field of Artificial Intelligence from the Computer Engineering Department, K. N. Toosi, University of Technology, Tehran, Iran. His current research interests include complex and social network analysis, machine learning and the use of machine learning and statistical methods in the analysis of networked data.



Seyed Hossein Khasteh received the B.Sc. degree in electrical engineering, the M.Sc. degree in Artificial Intelligence and the Ph.D. degree in Artificial Intelligence all from the Sharif University of Technology, Tehran, Iran. He is currently an assistant Professor with the Computer Engineering Department, K. N. Toosi, University of Technology, Tehran, Iran. His current research interests include social network analysis, machine learning and big data analysis.