International Journal of Computational and Electronic Aspects in Engineering
Vol. 5, Issue 4, Dec 2024, pp 141 - 147
https://doi.org/10.26706/ijceae.5.4.20241102

Method for Mining the Opinion Leaders in Social Networks

Mustafa K. Alasadi

Faculty of Computer Science and Information Technology, University of Sumer, Al-Rifai 64005, Thi Qar, Iraq

***Correspondence:** mustafa.kamil@uos.edu.iq

Abstract: Social network is one of the most prominent phenomena that greatly influence life in the present time. These media have changed the way people interact with others, the ways they communicate, and participate in global and local events. individuals on social network platform greatly influence each other's opinions and trends, but the degree of influence varies from one person to another greatly depending on many factors, such as fame, published content, target audience, and interaction with followers. This leads to varying influence of individuals on these platforms. Influential people can be opinion leaders on social platforms. In fact, these leaders constitute a large part of the process that influences the decisions of many individuals and companies, whether at the level of product selection or cultural and social trends. Discovering and identifying opinion leaders is very important for several reasons: directing public opinion, marketing and advertising, political and social analysis, and others. In this paper, we proposed a method to discover opinion leaders by focusing on the structure of social graphs and the position occupied by individuals. The centrality of nodes representing individuals were calculated by utilizing two essential measures: the degree centrality and betweenness centrality, but with different weights that were determined in a studied way. The method was tested on real-world dataset for three social networks and the results were promising when compared with the baseline.

Keywords: social network; Opinion Leaders; Degree Centrality; Betweenness Centrality.

Article – Peer Reviewed
Received: 5 Sept 2024
Accepted: 11 Nov 2024
Published: 3 Dec 2024

Cite this article: Mustafa K. Alasadi, "Method for Mining the Opinion Leaders in Social Networks", *International Journal of Computational and Electronic Aspects in Engineering*, RAME Publishers, vol. 5, issue 4, pp. 141-147, 2024.
https://doi.org/10.26706/ijceae.5.4.20241102

1. Introduction

Social network has become a crucial component of our everyday lives and plays a significant role in various personal, social, economic, and cultural dimensions [1][2]. Platforms like Facebook and Twitter offer a means to connect with friends and family, enabling individuals to engage with others who share similar interests or professions, thereby increasing opportunities for social interaction and creating new networks of relationships [3][4]. These communication platforms have evolved into a primary source for breaking news and information, allowing people to stay updated on global and local events instantly and share their thoughts and ideas regarding political and social matters [5].

In the field of social network analysis, social networks are typically represented as graphs $G\ (V, E)$ which consist of nodes V that symbolize individuals or entities within the network, and set of edges E that represent the relationships or interactions among these individuals, where edges may be directed or undirected based on the type of relationship present on social platform [6][7].

Individuals establish connections with one another in social networks through various relationships, and the dynamics of these networks influence how quickly and effectively content spreads. Individuals in these networks differ in their ability to influence others, for instance, individuals with a vast number of connections often wield greater influence [8][9]. This category includes celebrities, influencers, and journalists who boast millions of followers.

Social network opinion leaders are those individuals who hold significant influence over their communities or online audiences [10]. The concept of opinion leadership emerged in the field of communication studies during the 1940s and 1950s. In 1944, Professor Lazarsfeld discovered through his research that public communication does not directly reach the masses; instead, it is first filtered by opinion leaders before it reaches the general population [9][11].

Opinion leaders are the individuals greatly affect their communities or online audiences. They typically excel in specific domains such as technology, health, and politics and have numerous followers who place trust in their views and consider their recommendations, because they have strong communication skills that enable them to build relationships with their followers and provide influential content [12][13].

Identifying opinion leaders in social network is crucial across various fields, including marketing, politics, education, and social influence, due to their considerable capacity to sway the opinions and attitudes of followers and shape their thoughts and decisions based on the information these leaders share.From a marketing and brand promotion perspective, social media opinion leaders serve as powerful assets in marketing products and services [10][14]. They possess the ability to affect purchasing choices and draw attention to specific brands, thereby boosting sales and enhancing popularity.

Regarding social issue awareness, opinion leaders can help disseminate information concerning social and humanitarian issues that matter to society, such as environmental challenges, human rights, and equality [15][16]. They can significantly motivate individuals to engage with these matters and contribute to effecting change. In summary, recognizing opinion leaders on social media enhances the understanding of the digital community and aids in influencing both individual and collective decisions through their capacity to shape public sentiment.

In social network platforms, numerous opinion leaders can be found, but pinpointing a reliable and genuine opinion leader is quite a challenging endeavor [12]. Numerous studies have been conducted to identify the opinion leader within a social network. The researchers investigated different methods based on a variety of factors such as “trust models and metrics, the total number of tweets, fuzzy logic, belief network, game-theoretic model, clustering, and text mining and many more” [17][18] [19]. In this paper, we propose a technique for recognizing opinion leaders based on the centrality scores of the nodes within a social network, specifically focusing on both degree centrality and betweenness centrality but with varying weights. Since the centrality score of nodes is a vital measure of individuals' ability to disseminate content and attain greater influence through their connections. We performed an experiment and simulation involving three actual social networks. The outcomes were documented, analyzed, and debated. The findings indicated that the suggested approach for identifying opinion leaders is a promising and dependable method.

2. The Proposed Methodology

This essential section of the study, explains the proposed methods and techniques that were used to collect and analyze data in order to achieve the research objectives. The methodology explains how the research was conducted and why these specific methods were chosen. Figure 1 shows the framework of mining the opinion leaders.

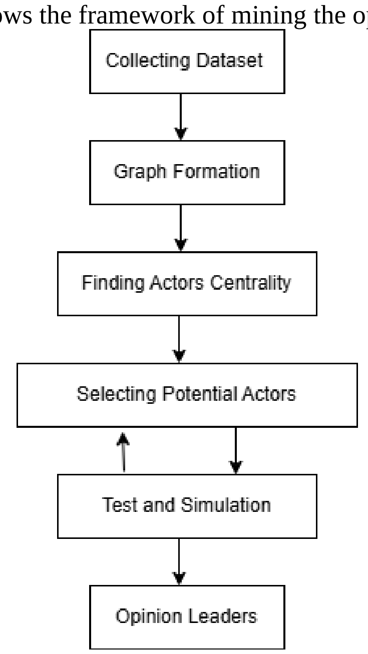

Figure 1. The general structure of proposed method

2.1 The Datasets

The Datasets considered one of the basic tools in scientific research in various fields; as it represents the basic sources that researchers rely on to obtain results and analyses. A dataset can be defined as a set of data that has been collected or created for the purpose of conducting a specific analysis or study. In this work four different types of datasets were carefully selected to suit the paper objective in order to reach more accurate and reliable conclusions. Below is an explanation of the datasets used in the experiment, along with an explanation of the nature of the information contained therein. Table 1 provides an overview of the fundamental statistics of the used data, such as the number of nodes and edges, in addition to mentioning the type of relationship between individuals in the social network, as well as the references of obtaining the data.

Table 1. The datasets overview

Name	No. nodes	No. edges	Type	Ref.
Twitter	475	13,289	directed	[20]
Facebook	4039	88234	undirected	[21]
Sina Weibo	320	526	directed	[22]

i- Twitter

This network illustrates the Twitter interaction dynamics among the 117th United States Congress, encompassing both the House and Senate. The foundational data was gathered using Twitter's API, and then the actual transmission probabilities were measured based on the proportion of instances in which one member re-tweeted, quoted, replied to, or referenced another member's tweet [20].

ii- Facebook

This dataset contains "circles or friends lists from Facebook" [21]. The data has been anonymized by substituting the original internal user IDs with new values. Furthermore, while the dataset includes feature vectors, their meanings have been intentionally obscured. For example, where the original dataset might have listed a feature as "political=Democratic Party," the revised dataset simply presents it as "political=anonymized feature 1." Hence, even though the anonymized data allows for determining whether two users share the same political views, it does not reveal the specific political affiliations of each user. Although we did not utilize the user's profile in the experiment, it is an important feature that can be used in future research because of its great information.

iii- Sina Weibo

Sina Weibo [22], commonly referred to as Weibo, is a Chinese microblogging platform that resembles Twitter. It serves as a media for fostering user connections, enabling the sharing, dissemination, and reception of information. Users can upload images and videos for public sharing through both the website and mobile application. Other users are encouraged to engage through comments, which can include text, pictures, and videos, as well as a multimedia instant messaging service. Sina Weibo has emerged as a primary source for a diverse array of information, including breaking news, social events, and product insights. Understanding user interests and behaviors is of significant value, as it presents opportunities to enhance comprehension of information dissemination mechanisms and identify opinion leaders within social network sites. User interaction is pervasive on the platform; users can leave comments on others' posts and respond to comments on various threads. Additionally, Sina Weibo promotes user participation in its identity verification program to further enhance the integrity of the platform.

2.2 Graph Formation

For each data used in the experiment, a graph was created that represents the nature of the relationships between social network users. As a result of this step, three graphs were created. Representing social networks through a collection of nodes and edges is a widely utilized methodology for comprehending and visualizing the intricate relationships among various individuals or entities within a network. This approach is commonly referred to as graph representation or graph network representation G(V, E). In this framework, each individual or entity in the network is denoted by a node, V = (v1, v2,...,vn) while the relationships among those entities are indicated by edges E = (e1, e2,..., en). Nodes: Each node signifies a distinct entity within the network, which may encompass persons, groups, companies, or even websites. In the context of social networks, nodes predominantly represent individuals. Edges: An edge serves as a connector between two nodes, illustrating the relationship that exists between the entities. Within social networks, edges can embody different types of relationships, such as friendships, followers, messages, or other forms of interactions. Furthermore, edges may be classified as directed—signifying a one-way relationship, such as a follower on social media—or undirected—indicating a reciprocal relationship between the two nodes, as seen in friendships.

2.3 Finding Nodes Centrality

Determining the centrality of actors within social networks involves evaluating significance of each actor (or "node") based on their relationships with other actors. In the context of social networks, centrality pertains to the capacity of an individual or entity to cause the dissemination of information, make decisions, or impact the overall network. There are various measures available for centrality in social networks, here we utilize two metrics: Betweenness and degree centrality.

Betweenness Centrality: This centrality measures how much information passes through a particular node. A node that is between many other pairs of nodes may have considerable influence within a network by virtue of their control over information passing between others. The Betweenness of any given node is calculated as in equation 1 [23].

$$Cb(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (1)$$

Where " σ_{st} is the total number of shortest paths from node S to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v"[23]. Figure 2 shows an example of how this measure is calculated.

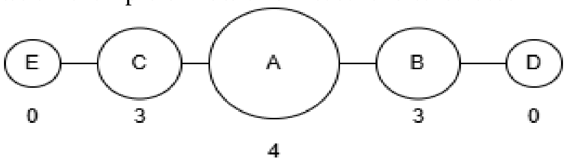

Figure 2. An example of node Betweenness [19].

Degree Centrality [24]: Degree centrality, the most foundational and conceptually straightforward measure, is defined as the total number of connections associated with a node (i.e., the quantity of ties a node possesses). This type of centrality is calculated by counting the number of links (relations) a node has. The more links, the more central the node. Equation 2 shows the calculation of this metric.

$$Cd(v) = \deg(v) \qquad (2)$$

$\deg(v)$ The number of links directly connected with node$v$.

2.4 Selecting Potential Actors

After calculating the centrality of the nodes in the graph, in this step of the proposed method we propose to rely on both measures Betweenness and degree centrality. We suggest relying on these measures differently by setting different weights to calculate the relative importance of each node in the graph, as illustrated in equation 3.

$$Imprtance\ (v) = \alpha Cb(v) + \beta Cd(v) \qquad (3)$$

The α, β are the weight associated with Betweenness and degree centrality score of the given node v . In this phase we utilize the Analytic Hierarchy Process to determine the value of α, β, which was proposed in [25] since it widely used in processing the complex decision-marking problems. Following the assessment of node importance, the nodes are systematically organized according to their significance, and the top k nodes are selected as potential opinion leaders.

2.5 Test and Simulation

To measure the effectiveness of proposed method, we utilize the Independent Cascade (IC) Propagation model to test and simulate the propagation process [26][27][28]. The model starts from a group of individuals who are exposed to a specific event, such as receiving a specific message or information. These individuals are considered as seed nodes. In the conducted experiment the top k potential opinion leaders specified in the previous step are used as seed nodes to study their influence on the other nodes in the graph. This model is based on the idea that information is transmitted between individuals (nodes) through links that connect them to each other. It is mainly used to understand how ideas spread and the dynamics of social diffusion, such as how advertising campaigns or fake news spread or in marketing campaigns through social networks. For evaluation purpose we conduct the experiment three times again by selecting the seed nodes randomly, then the top k Betweenness nodes, and finally top k degree centrality nodes.

2.6 Opinion Leaders

In the final step of the proposed method, the nodes that are considered opinion leaders are identified. These nodes are selected through the results of the influence simulation using the IC model from the previous step. The top N most influential nodes and those with the greatest ability to spread content throughout the social network are selected.

3. Results and Discussion

In this section, the results of the experiment applied according to the steps of the proposed method and the datasets used are reviewed. Figure 3 shows the graph obtained from the Sina Weibo data as an example of graph formation, where the data was represented on a graph in which the nodes appear in different colors according to the degree of centrality of the nodes.

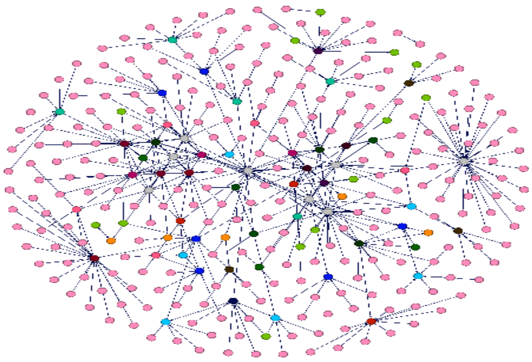

Figure 3. The generated graph of Sina Weibo dataset.

Table 2 presents the findings from the experiments carried out on three real-world social network datasets. The Independent cascade model was modified to be the cascade length of diffusion three steps only since it is efficient enough to obtain the findings. The experiment and the spread of the effect were conducted four times. First, the seed nodes were chosen randomly and the number of effected nodes were recorded in each dissemination hope. In the next trail, the seed nodes were chosen from the nodes with the highest degree centrality. Then the seed nodes were the node with the highest Betweenness. finally, the diffusion simulated with seed nodes that have highest importance calculated by applying the proposed equation 3 with α=0.7, β=0.3.

Table 2. The experiments results

Dataset	Cascade len.	Random	Degree	Betweenness	proposed method
Twitter	1	4	19	23	33
	2	9	33	61	76
	3	19	47	79	88
Facebook	1	40	192	232	282
	2	119	313	525	646
	3	161	481	678	702
Sina Weibo	1	3	12	16	22
	2	6	22	41	51
	3	12	32	48	60

4. Conclusion

By studying the results in the table 2, we notice that the number of activated or effected nodes is less when the seed nodes are randomly selected. In the Twitter data, we notice that the number of effected nodes in the first step of cascade is 4, after which 5additional nodes were activated in the next step, and finally in the third step, the total number of activated nodes was 19. when the seed nodes are selected from the nodes with highest scores of Degree and Betweenness centrality, the total number of activated nodes were 47 and 79, respectively.

The important thing is that the proposed method produced the highest number of effected nodes (88 in Twitter data, 702 in Facebook, and 60 in Sina Weibo) compared to the rest of the methods. This finding can be seen in the results of applying the experiment to the rest of the databases. From here, we conclude that relying on both Degree and Betweenness centrality in calculating the importance of nodes and choosing the nodes that lead the opinion leaders is a very effective method and can be generalized to be used on social networks.

References

[1] M. Macdonald, A. Gunderson, and K. Widner, "Exploring Interest Group Social Media Activity on Facebook and Twitter," *J. Quant. Descr. Digit. Media*, vol. 4, pp. 1–60, 2024, doi: 10.51685/jqd.2024.014.

[2] M. F. Mahdi, "Revolutionizing the Future Investigating the Role of Smart Devices In IOT," vol. 5, no. 1, pp. 1–15, 2024.

[3] P. Giulio, "Psychopathological profiles and trends of Italian social network users (Facebook, Instagram, Twitter, and TikTok)," *Ann. Psychiatry Treat.*, vol. 6, no. 1, pp. 053–061, 2022, doi: 10.17352/apt.000045.

[4] Rimpy, A. Dhankhar, and A. Dhankhar, "Sentimental analysis of social networks: A comprehensive review (2018-2023)," *Multidiscip. Rev.*, vol. 7, no. 7, pp. 1–33, 2024, doi: 10.31893/multirev.2024126.

[5] I. K. AI –Dulaimi, "The Use of Cloud Computing to Process Big Data: An Applied Study of the Virtual Library at the University of Mosul," *Int. J. Comput. Electron. Asp. Eng.*, vol. 4, no. 2, pp. 38–43, 2023, doi: 10.26706/ijceae.4.2.20239815.

[6] T. S. Hwere, H. Yakubu, and R. S. Isa, "Graph Models of Social Media Network As Applied to Facebook and Facebook Messenger Groups," vol. 9, no. 1, pp. 1–12, 2023, doi: 10.56201/ijcsmt.v9.no1.2023.pg1.12.

[7] C. Chouhan, S. Tiwari, and A. U. Rahman, "Social networks and representation of graph theory," pp. 3–6, 2023.

[8] M. K. Alasadi and G. I. Arb, "Community-based framework for influence maximization problem in social networks," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 3, pp. 1604–1609, 2021, doi: 10.11591/ijeecs.v24.i3.pp1604-1609.

[9] Mak and Vincent, "The Emergence of Opinion Leaders in Social Networks Vincent Mak," *Water*, no. 852, 2008.

[10] A. Aleahmad, P. Karisani, M. Rahgozar, and F. Oroumchian, "OLFinder: Finding opinion leaders in online social networks," *J. Inf. Sci.*, vol. 42, no. 5, pp. 659–674, 2016, doi: 10.1177/0165551515605217.

[11] S. Sciences, "Russian Journal of Agricultural and Socio-Economic Sciences, 8(20)," vol. 8, no. 20, pp. 20–26, 2003.

[12] M. Kang, T. Liang, B. Sun, and H. Y. Mao, "Detection of opinion leaders: Static vs. dynamic evaluation in online learning communities," *Heliyon*, vol. 9, no. 4, p. e14844, 2023, doi: 10.1016/j.heliyon.2023.e14844.

[13] M. K. Abbas, "Modelling WhatsApp Traffic Control Time-Based (WTCTB) for 5G Mobile Network," *Int. J. Comput. Electron. Asp. Eng.*, vol. 4, no. 4, pp. 110–118, 2023, doi: 10.26706/ijceae.4.4.20231003.

[14] A. M. Litterio, E. A. Nantes, J. M. Larrosa, and L. J. Gómez, "Marketing and social networks: a criterion for detecting opinion leaders," *Eur. J. Manag. Bus. Econ.*, vol. 26, no. 3, pp. 347–366, 2017, doi: 10.1108/ejmbe-10-2017-020.

[15] S. Walter and M. Brüggemann, "Opportunity makes opinion leaders: analyzing the role of first-hand information in opinion leadership in social media networks," *Inf. Commun. Soc.*, vol. 23, no. 2, pp. 267–287, 2020, doi: 10.1080/1369118X.2018.1500622.

[16] A. N. Ayesh, "Optimizing of Cloud Storage Performance by Using Enhanced Clustering Technology," vol. 5, no. 1, pp. 16–24, 2024.

[17] J. Turcotte, C. York, J. Irving, R. M. Scholl, and R. J. Pingree, "News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking," *J. Comput. Commun.*, vol. 20, no. 5, pp. 520–535, 2015, doi: 10.1111/jcc4.12127.

[18] F. Riquelme, P. Gonzalez-Cantergiani, D. Hans, R. Villarroel, and R. Munoz, "Identifying Opinion Leaders on Social Networks Through Milestones Definition," *IEEE Access*, vol. 7, pp. 75670–75677, 2019, doi: 10.1109/ACCESS.2019.2922155.

[19] M. M. Madbouly, S. M. Darwish, and R. Essameldin, "Modified fuzzy sentiment analysis approach based on user ranking suitable for online social networks," *IET Softw.*, vol. 14, no. 3, pp. 300–307, 2020, doi: 10.1049/iet-sen.2019.0054.

[20] C. G. Fink *et al.*, "A centrality measure for quantifying spread on weighted, directed networks," *Phys. A Stat. Mech. its Appl.*, vol. 626, 2023, doi: 10.1016/j.physa.2023.129083.

[21] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 539–547, 2012.

[22] X. Cui and H. Shi, "A Comprehensive Study on Data Extraction in SINA WEIBO," *Int. J. Artif. Intell. Appl.*, vol. 7, no. 4, pp. 47–57, 2016, doi: 10.5121/ijaia.2016.7404.

[23] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," vol. 132, no. Msam, pp. 300–303, 2017, doi: 10.2991/msam-17.2017.68.

[24] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, "Opinion leader detection: A methodological review," *Expert Syst. Appl.*, vol. 115, pp. 200–222, 2019, doi: 10.1016/j.eswa.2018.07.069.

[25] T. L. Saaty, "Basic Theory of the Analytic Hierarchy Process : How To Make a Decision," *Rev. la Real Acad. Ciencias Exactas, Físicas y Nat.*, vol. 93, no. JANUARY 1999, pp. 395–423, 1999.

[26] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *SIGMOD Rec.*, vol. 42, no. 2, pp. 17–28, 2013, doi: 10.1145/2503792.2503797.

[27] M. K. Alasadi and H. N. Almamory, "Diffusion model based on shared friends-aware independent cascade," *J. Phys. Conf. Ser.*, vol. 1294, no. 4, 2019, doi: 10.1088/1742-6596/1294/4/042006.

[28] E. A. Abbas and H. N. Nawaf, "Influence maximization based on a non-dominated sorting genetic algorithm," *Karbala Int. J. Mod. Sci.*, vol. 7, no. 2, pp. 139–150, 2021, doi: 10.33640/2405-609X.2891.