Central European Journal of Operations Research (2022) 30:367–394
https://doi.org/10.1007/s10100-021-00738-5

Comprehensive decomposition optimization method for locating key sets of commenters spreading conspiracy theory in complex social networks

Mustafa Alassad[1] · Muhammad Nihal Hussain[1] · Nitin Agarwal[1]

Accepted: 25 January 2021 / Published online: 7 February 2021

Abstract

With the power of social media being harnessed to coordinate events and revolutions across the globe, it is important to identify the key sets of individuals that have the power to mobilize crowds. These key sets have higher resources at their disposal and can regulate the flow of information in social networks. They can maximize information spread and influence/manipulate crowds when they are coordinating. But due to the inherent drawbacks in node-based and network-based community detection algorithms, neither of these types of algorithms can be used to detect/identify these key sets. In this study, we present a bi-level max-max optimization approach to identify these key sets, where the degree centrality is used to identify individuals' influence at the commenter-level, while the network-level is designed to evaluate the spectral modularity values. We also present a set of evaluation metrics that can be used to rank these key sets for an in-depth investigation. We demonstrated the efficacy of the proposed model by identifying key sets hidden in a YouTube network spreading fake news about the conflict in South China Sea. The network consisted of 47,265 comments, 8477 commenters, and 5095 videos. A co-commenter network was constructed, where two commenters were linked together if they comment on same video. The proposed model efficiently identified key sets of commenters spread information to the whole network to manipulate YouTube's recommendation and search algorithm to increase the information dissemination. Moreover, the projected approach could identify sets of commenters that were key connectors to multiple groups, high influence across the network, higher interactions, and reachability than other regular communities. Besides, the Girvan–Newman modularity method, the depth-first search method, and text analysis was applied to validate the outcomes, categorize the identified key sets, and monitor the commenters' behaviors and information spread strategies in the network. In addition, the model considered a multi-criteria problem to rank these key sets of commenters based on the small real-world networks' features.

Extended author information available on the last page of the article

Keywords Social network analysis · Focal structure analysis · Fake news spreaders · Bi-level optimization · Multi-criteria optimization method · Misinformation · Disinformation

1 Introduction

Social media services have changed our world rapidly due to social-networking functionalities that allow millions of people to share and track information or live news stream, daily, and hourly. Social media platforms such as Facebook, Twitter, and Telegram acting like an important, free, and easy communication tools to help people find common interests and share their attentions with others. However, in the recent years these tools have been mis-used by deviant users to influence other users for political gains, provoke anti-government protest, spread fake news, and spread mis-leading information to destabilize societies. Likewise, such kind of malicious users, used social media platforms as powerful tools to mobilize crowds in many recent big anti-government movements like the "Egyptian Revolution", "Yellow Vests Movement", "Hong Kong protest", "The Arab Spring", and most recent one is the "Iraqi anti-corruption protest". Where most of these movements were orchestrated by users on Facebook, WhatsApp, YouTube, and Telegram (Şen et al. 2016).

The rise in malicious behaviors, prevalent the misinformation on social media platforms and their perceived influence on politics, economic, health systems, and many other important parts of our society. And due to huge damages, these topics attracted many researchers, academic institutions, and other administrations' funding resources to study and counter these behaviors. These activities have also compelled the government agencies and academic institutions to collaborate and fight against such behaviors. The resultant collaborations and studies, they uncovered solutions to several questions such as: Who are the influential users on social networks? How can we stop users' influence? How do we identify groups of leaders acting in different parts of the network? How can we elaborate models' outcomes on social networks? and many other related questions. Likewise, the proposed model in this paper will digest more into these questions, provides the reader intuitive observation, scientific, quantitative and qualitative comprehended outcomes.

This research focuses on identifying key sets (influential sets, focal structure, or central groups) of users who are able to influence the maximum number of users in different parts of the network. Additionally, these influential sets were remained hidden in the network and the regular community detection methods failed to identify them due to gap in the methods' features and analysis.

To overcome the shortcomings in the user-level community detection methods such as centrality methods (Zafarani et al. 2014) and the group-level community detection methods like the modularity method (Girvan and Newman 2002a); the model proposed here combines these two well-known community detection methods (centrality method and modularity method) to bridge the gap between the user-level's analysis and the group-level's analysis in graph theory (Zafarani et al. 2014), as displayed in Fig. 1. The resultant combination is a bi-level linear optimization problem to realize/observe the interactions between the two levels' considerations as

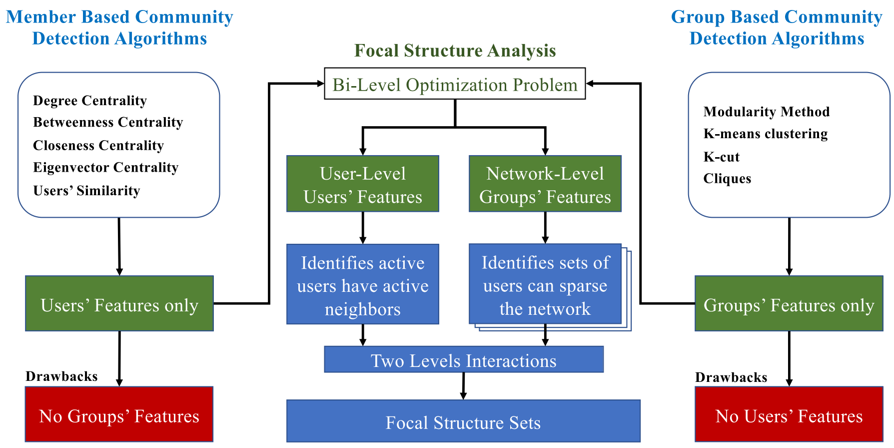


Fig. 1 Focal structure analysis, a bi-level maximization network model, i.e., identifying authoritative individuals and identifying communities

presented in Fig. 2. The model will utilize any connected unimodular network and measure the user-level's features such as the degree centrality and clustering coefficient values to find the active users linked to set of active neighbors; this will help identifying the active local communities (Zafarani et al. 2014). Later, the model uses the group-level analysis, where the spectral modularity method is utilized to measure the impact generated by the local communities on the entire network. Both levels "user-level and group-level" will maximize their solutions to find the central sets of users that can maximize the network's influence; including active users able to influence the maximum number of users in the network and maximize the network's sparsity.

In addition, to supplement the model's analysis and validate the outcomes, the solution procedure will implement a set of real-world small network metrics (Zafarani et al. 2014). For this purpose, the model utilizes a multi-criteria optimization problem to rank and confirm the sets were not randomly selected from the network as explained in Sect. 5.

Moreover, to demonstrate the efficiency of the proposed model, we applied the model into a network of commenters on a YouTube channel that had more than 15 million views spreading fake news and conspiracy theory videos related to the conflict in South China Sea. The dataset was curated using the methodology explained by Hussain et al. (Hussain et al. 2018). Finally, the text analysis is employed to study the sets commenters' behaviors and actions then categorize them into different categories.

The rest of the paper is organized as follows. First, a selection of the existing literature on influencers and community detection algorithms are reviewed in Sect. 2. Next, the empirical study is described in Sect. 3, and Sect. 4 discusses the model's complexity analysis utilizing a toy example. Experimental results are in Sect. 5. Section 6 is to validate the results. Lastly, we discussed our conclusions, limitations, and ideas for the future research work in Sect. 7.

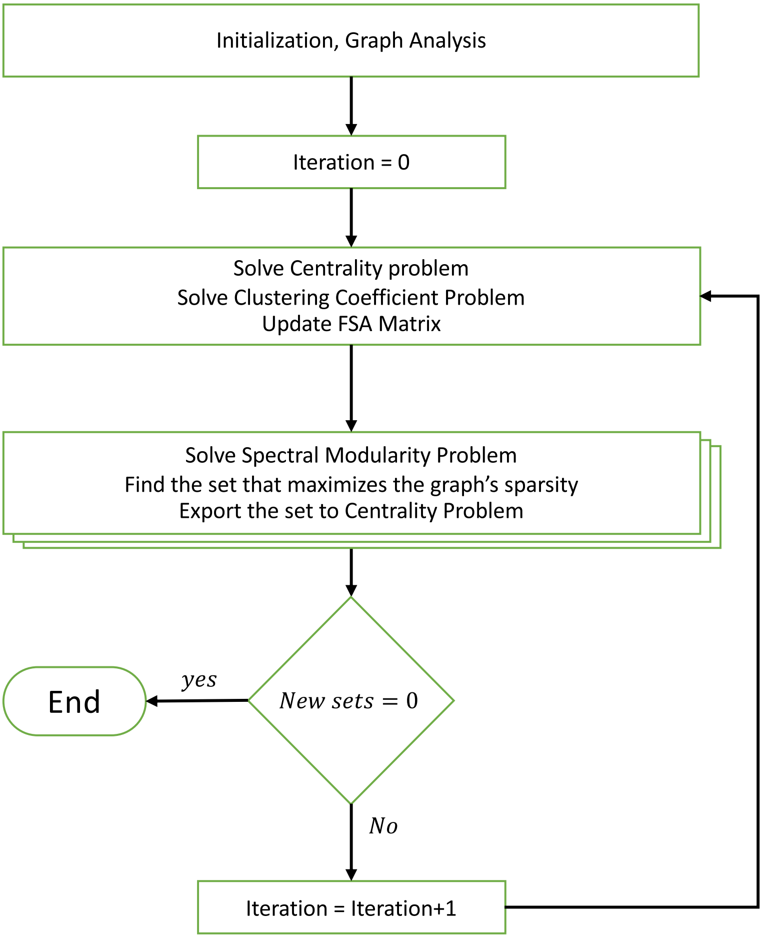


Fig. 2 Overall bi-level model structutre

2 Related works

Complex network analysis has received larger focus in recent years. There has been a greater effort within the academia institutions to simplify the analysis, linearize the implemented methods, and optimize the overall solutions. Complex fake news analysis in social media gained more attention after the 2016 US presidential elections (Horne and Adali 2017) where many datasets from different platforms were collected and investigated. Major number of these studies were abstracted to find the online active users who were responsible for information dissemination and other related conspiracy theories. Many researchers applied traditional graph theories to cluster the influential users (Dinh et al. 2016), how to sparse complex social networks into smaller communities to simplify the analysis (Chen et al. 2017). Methods such as centrality (Zafarani et al. 2014), PageRank (Page et al. 1998), and HITS

(Kleinberg 1999), produced incredible results with respect to the users' aspects only. Other researchers such as Herzig et al. (2014) and Yang et al. (2016) investigated the users' ability to disseminate information. Briscoe et al. (2013) studied the truth of news from unidentified users in online platforms. However, in an advanced research proposed by You et al. (2020), a three-stage algorithm implemented to detect communities based on the users' local and the global information spread. In this research, the authors projected methods to identify the central nodes, utilized the label propagation in the second stage, and the third stage considered the communities combination. In general, the mentioned methods can project the influence generated by any online users based on the available resources, the spread of false information and conspiracy theories, and the identified authoritative users and their followers in the social networks.

However, since these methods solely depends on the user's aspect in the analysis; we believe that only one influential user cannot amplify the spread of information to thousands of users and mobilize tens of thousands of people (Şen et al. (2016)) on social media. Şen et al. (2016) claimed that there should be sets of like-minded users coordinating to mobilize crowd, share data, post movement's schedules, and spread information to influence the maximum number of users in the network.

On the other hand, many researchers digested into the group-level analysis to simplify the complex network analysis. Many methods such as modularity proposed by Newman (2006) have been extensively used to cluster complex networks into smoother communities. However, this method is an NP-hard problem (Chen and Wang 2009) and lots of optimization problems were implemented to optimize the outcomes. Hu and Liu (2016) applied a revolutionary inverse modeling methods to partition the complex communities in large networks, they utilized modeling based on the multi-objective method. Leskovec et al. (2007a) investigated the social networks' behaviors to find patterns of cascading behavior in large blog graphs. The local edges centrality method was proposed to detect communities in large networks by utilizing the modularity method and the division method as explained in Hu and Liu (2016). A convex formulation was proposed by Chan and Yeung (2011) to optimize the patterns clustering in complex networks. Others such as Alinezhad et al. (2020) proposed a two novel mathematical programming approach to integrate the topological structure and the nodes similarities in complex networks. Waltman and Van Eck (2013) applied a smart local moving algorithm for detecting communities in large scale networks. A cutting plane algorithm with heuristics approaches was recommended by Izunaga and Yamamoto (2017) to help with the communities separation problem. Sato and Izunaga (2017) maximized the graph separation by using a branch-and-price method with MILP formulation. Finally, a novel hierarchical clustering algorithm advised by Xie et al. (2020) to measure the distance between two users then grouped them into one cluster.

The other major scope is the users' interactions and the information exchange between online influential users on social network. Jones and O'Neill (2010) proposed an intresting reseach about various relationship aspects such as the similarities between different users in the network, where they compared the users' entities to the sets identified by their algorithm. Likewise, users' behaviors in complex networks investigated by Leskovec et al. (2007b) and Li et al. (2018), where the proposed work stemmed into identifying the influential users based on their neighbors' activities in complex networks.

However, there is no objection we can raise to all mentioned outstanding methods and their outcomes, but their analysis did not consider the influential users' actions and the influential groups' impacts on the users and the network in whole (Şen et al. 2016). For example, implementing solely traditional community detection methods into the social networks would result in small groups or without noticeable impression on the information flow, conspiracy theories, nor including central influential users in the outcomes (Şen et al. 2016). Such limits would stretch the gap in the analysis and will cause the influential sets of users remain hidden inside the big communities (Alassad et al. 2019a).

The objective of this research is to integrate the traditional community detection algorithms to overcome these shortcomings and identify coordinating groups of users able to spread and influence the maximum number of users in complex social network. However, in this battleground, Şen et al. (2016) proposed the state of the art model, they presented a greedy algorithm to identify the key set of user called (focal structure sets) in a Facebook network. Alassad et al. (2019a) presented the degree centrality and the modularity methods together to optimize the focal structure analysis and to overcome Şen et al. (2016) model's drawbacks. Alassad et al. (2019a) proposed a model to identify the hidden key sets of co-commenters in a YouTube network. In addition, Alassad et al. (2020) investigated the amplification and coordination activities of influential sets of users in social networks. Also, Alassad et al. (2019b) proposed a related research to finding the fake news key spreaders in complex social networks by using bi-level decomposition optimization method.

For this study, we extended the previous research proposed in Alassad et al. (2019b). The proposed research here invested significant efforts to include the model's complexity analysis, study the network resiliency, analyze the sets' impacts on the entire network, and categorize the identified sets into different levels based on the sets' and users' behaviors, comments, and their interactions with other users.

3 Proposed methodology

To overcome the shortcoming in the traditional community detection methods (Zafarani et al. 2014; Girvan and Newman 2002a) and the drawbacks mentioned in the state of the art model (Şen et al. 2016), the model implements two levels of analysis followed by two verification levels as explained in this section.

3.1 Problem definition

An undirected social network on YouTube is spreading conspiracy theory, $G = (V, E)$, where V represents the commenters in the channel, and E represents relationships between the commenters, i.e. if two users comments on the same posted video on the YouTube channel. Using the bi-level problem, identify the unknown smallest possible κ intensive sets of V users that can influence the maximum number of users in G, we refer to as Focal Structure Sets.

3.2 Data set

3.2.1 Karate club network

This network considered into this research as a toy example to demonstrate the model's complexity analysis. The implemented network was presented by Wayne W. Zachary from 1970 to 1972, it captures 34 members of a karate club as shown in Fig. 3 (left side). The network reflects the disagreement in a karate school, where the administrator "John A", and the instructor "Mr. Hi" went into a conflict, which resulted to split the club into two separate schools (Zachary 1977).

3.2.2 YouTube video channel

This dataset was collected from a YouTube channel spreading conspiracy theory videos related to the conflicts in South China Sea (Alassad et al. 2019a). The undirected graph presented in Fig. 3 (right side), was derived from the dataset. This network includes 8477 commenters posted their comments on 5095 videos. The users' conspiracy theories increased the network's interactions into more than 1 million edges.

Also, the network shown in Fig. 4 represents a filtered network retaining only edges that had weight of ten or more. The network clustered into different communities via the modularity method. The network shows few central users as ego centers and big complex communities which makes it hard to track the spread of information and the conspiracy theories between the channel's users.

3.3 Commenter-level sphere of influence

The degree centrality method utilized in this research helped to measure the commenter's power, number of neighbors linked to each commenter, and their capacities

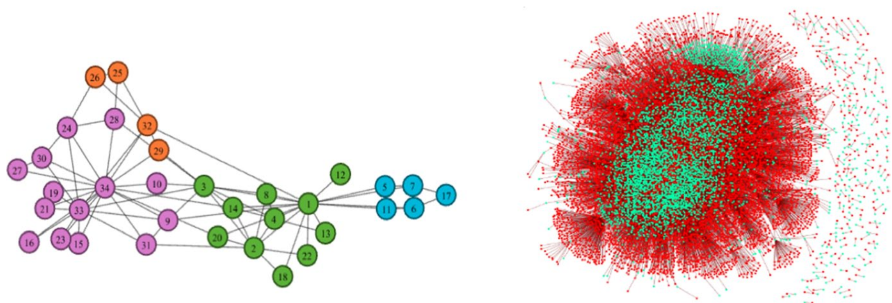


Fig. 3 Left side, the Karate Club network. Right side, the YouTube commenter network (the videos are in green color and the commenters in red color) (color figure online)

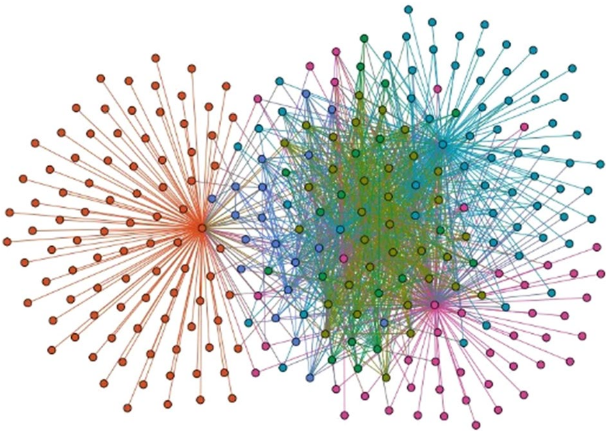


Fig. 4 Weight ten YouTube commenters networks

to spread information to other users (Tsung et al. 2016). In addition, implementing this well-known centrality method in the model would help to observe the most active commenters in the network as shown in Fig. 5. The outcomes illustrate that the most central commenters are among the top 50 commenters identified in this level of the analysis. In other words, these top 50 users posted higher number of comments into the YouTube channel, linked to higher number of users, and occupied central positions in the network's structure.

Next step is to investigate the commenters' neighbors and explore which commenters did build strong coordinating connections with other active commenters. To find the active neighbors, Clustering Coefficient Method (Zafarani et al. 2014) is employed to measure the connections between the commenter's neighbors as revealed in Fig. 6. The outcomes would provide an essential information about each commenter's neighbors, indicating that such commenters actually built coordinating communities to increase

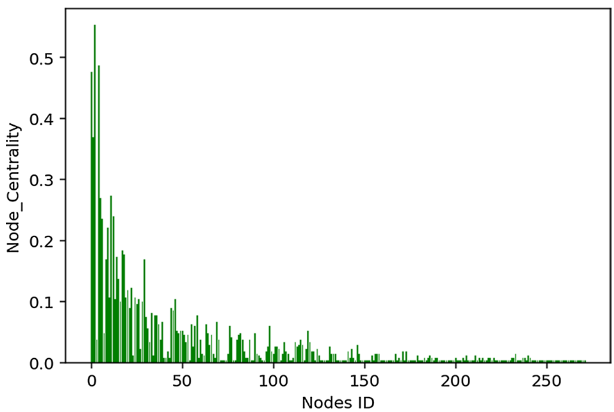


Fig. 5 Users' degree centrality values

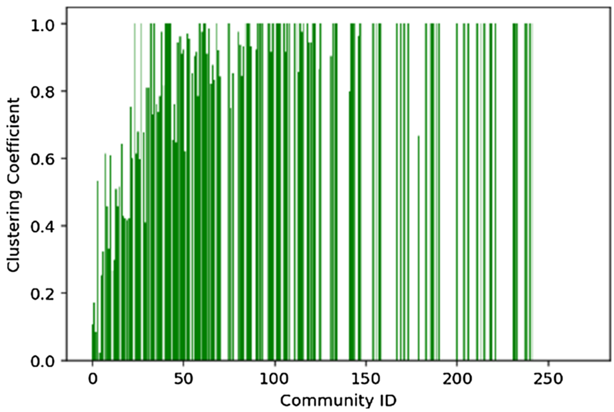


Fig. 6 Users' neighbors average clustering coefficient values

their organization level and to maximize their influence in the network. This coordination will allow the members to communicate very frequently and will enable them to control the information flow from/to other parts of the network.

The model at the user-level (or commenter-level) considers the commenters' features only. To implement the objective function and maximize the results, the model utilized Eqs. 1–12 to optimize the degree centrality values and the overall solution procedure. These equations are designed to maximize the commenters' degree centrality values to find the central active commenters connected to active neighbors.

The results from the commenter-level will nominate the non-dominated active commenters liked to active users as local communities. After that, when the model finishes the analysis in this level, the group-level will get the results to measure their influence in the network as a whole. The commenter-level equations are presented as follows.

$$max \sum_{i=1}^{n} \delta_i \tag{1}$$

Subject to

$$\delta_i = \left\{ d\vec{c}_1 \leqslant d\vec{c}_2 \leqslant d\vec{c}_3 \leqslant \cdots \leqslant d\vec{c}_i \right\} - \overline{dc_j^Q} \quad \forall\, i,j \tag{2}$$

$$d_i^c = \sum_j m_{ij} \quad \forall\, i \tag{3}$$

$$d_i^c \geq 2 \quad \forall\, i \tag{4}$$

$$D_G^L = \frac{1}{n} \sum_{i=1}^{n} d_i^c \tag{5}$$

$$D_G^L < d_i^c \leq D_G^U \quad \forall i \tag{6}$$

$$a_i^c = \frac{(\#\, of\ triangles) \times 3}{\#\, of\ connected\ triples\ of\ nodes} \quad \forall i \tag{7}$$

$$AC_G^L = \frac{1}{n} \sum_{i=1}^{n} a_i^c \tag{8}$$

$$AC_G^L < a_i^c \leq AC_G^U \quad \forall i \tag{9}$$

$$\overrightarrow{C_v} = \{\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{c_3}, \ldots, \overrightarrow{c_i}\} - \overline{c_j^Q} \quad \forall i,j \tag{10}$$

$$\overline{\vec{c}\delta_i}_{\,n \times k} = \overrightarrow{C_{\delta i}} \quad \forall i \tag{11}$$

$$F = \left\{ c_0, \overline{c_j^Q}, \overline{c_{j+1}^Q}, \ldots, \overline{c_\kappa^Q} \right\} \quad \forall j, k \tag{12}$$

Equations (1) and (2) are employed to maximize the commenters' degree centrality values (δ_i), where $i = 1$ is the least central commenter in the network and $i = n$ is the most influential commenter. The parameter $\overline{dc_j^Q}$ in the second part of this equation is used to transfer information from the group-level and help the model to exclude commenters from the solution procedure if they were identified in the group-level as shown in Fig. 8. Equation (3) is used to calculate the commenter's i influence in the network, called d_i^c, and m_{ij} is the sum of neighbors j linked to commenter i. Constraint (4) is to make sure the selected commenters have more than one neighbor. Equation (5) is to generate the network's degree centrality lower bound D_G^L. Equation (6) is to make sure the selected commenters are satisfying the degree centrality bounds, where D_G^U is the upper bound. Equation (7) is used to measure the commenter's neighbors' activities, or their clustering coefficient values. For example, if $a_i^c = 1$, then the commenter i has active neighbors or fully connected commenters and neighbors, where $a_i^c = 1$ is the clustering coefficient's upper bound UB, otherwise if $a_i^c = 0$, then commenter's i neighbors are not communicating/coordinating (star shape/chain set). The model will exclude such commenters from the solution procedure. Equation (8) is to calculate the network's average clustering coefficient value, where AC_G^L is the lower bound LB and $UB = 1$ is the highest possible value. Equation (9) is to measures the commenter's i neighbors' activities a_i^c and select commenters linked to active neighbors that satisfying the limits in both bounds (UB, LB). In Eq. (10), the vector representation of the sorted group of commenters and their neighbors, and the second part is a special parameter $\overline{c_j^Q}$ used to transfer information from the group-level to other level. Equation (11) is used to transfer information from the commenter-level to group-level utilizing the special parameter $\vec{c}\delta_{i n\times k}$, where it is a binary vector parameter consisted of n rows (number of commenters in the network) and k columns is the number of selected active set of commenters from the commenter-level. Equation (12) is the set of the active commenters linked to active neighbors that maximized the objective functions in both levels after many iterations. These sets are the final outcomes called Focal Structure Sets.

In summary, the model in this section will optimize the commenters' centrality values to find the most active commenters considering their neighbors' connectivity into the solution procedure. The identified influential commenters linked to active neighbors would generate local intensive sets of commenters (or focal structure sets); this means that these commenters have enough resources to communicate, coordinate, and spread information to other users in the network.

Moreover, the analysis at this level is not enough, the next step is designed to maximize the influence generated by each of these active sets on the entire network. For this purpose, the identified sets of commenters from the commenter-level are further processed at the group-level as presented in Fig. 8.

3.4 Group-level analysis

The objective in this section is to maximize the network's sparsity values by feeding the influential sets of commenters identified in the commenter-level into the modularity method. To begin such step, the vector parameter $\vec{c}\delta_{in\times k}$ will transfer the local sets identified in the commenter-level into the group-level, then the modularity method will measure the changes in the network sparsity.

Moreover, the aim of using the spectral modularity method (Zafarani et al. 2014; Girvan and Newman 2002a; Newman 2004a, b), is to escalate the analysis from the individual's influence to the global influence. Likewise, this step is to observe the impact produced by each active set of commenters and their active neighbors to other users and the entire network.

The objective is to maximize the spectral modularity values ϱ_j presented in Eq. (18) by utilizing the transferred optimized solutions from the commenter-level via parameter $\vec{c}\delta_{in\times k}$ iteratively. The model used two vector parameters to exchange information with the commenter-level, the parameter $\overline{c_j^Q}$ is to transfer the best sets of commenters that maximized ϱ_j, and $\overline{dc_j^Q}$ is the degree centrality for the set of commenters in $\overline{c_j^Q}$, as presented in Figs. 7 and 8. Likewise, in this level, the model will only use parameter $\vec{c}\delta_{in\times k}$ to transfer the central commenters from the commenter-level then introduce them into the spectral modularity equation as shown in Eq. (15).

To achieve this, Eqs. (13)–(20) are designed to maximize the group-level's objective function and exchange information with commenter-level as follows.

$$max \sum_{j=1}^{n} \varrho_j^M \tag{13}$$

Subject to

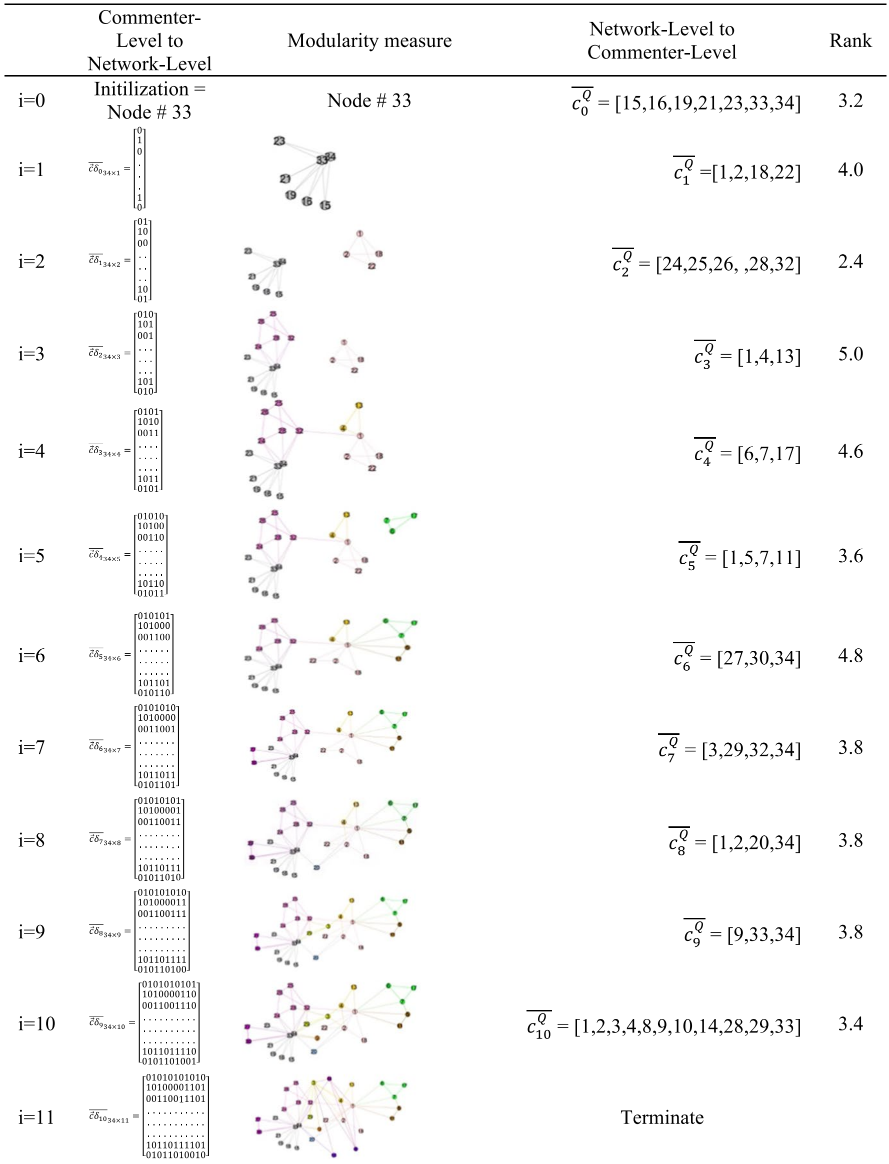


Fig. 7 Model's iterations to find intensive sets of members in Karate club network

$$\varrho_j^M = \{\varrho_1, \varrho_2, \varrho_3, \dots, \varrho_j\} \quad \forall j \tag{14}$$

$$\delta_j = \{\vec{c_1}, \vec{c_2}, \vec{c_3}, \dots, \vec{c_n}\} - \overline{\vec{c}\delta}_{in\times k} \quad \forall i, j \tag{15}$$

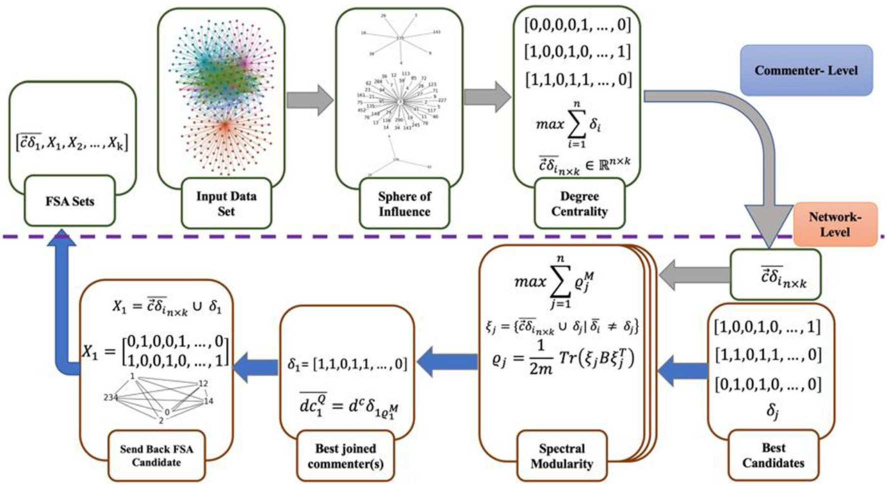


Fig. 8 Interactions between commenter-level and group-level, where the assigned parameter $\left(\overline{\vec{c}\delta_{i_{n\times k}}}\right)$ transfers information from the commenter-level to the group-level. Likewise, the group-level will transfer the best set of users that jointly maximized the spectral modularity values $\left(\overline{c_j^Q}\right)$ into the commenter-level

$$B = A_{ij} - \frac{dd^T}{2g} \quad \forall\, i,j \tag{16}$$

$$\xi_j = \{\overline{\vec{c}\delta_i} \cup \delta_j | \overline{\vec{c}\delta_i} \neq \delta_j\} \quad \forall\, i,j \tag{17}$$

$$\varrho_j = \frac{1}{2m} Tr\left(\xi_j B \xi_j^T\right) \quad \forall\, j \tag{18}$$

$$\varrho^l \leq \varrho_j \leq \varrho^U \quad \forall\, j \tag{19}$$

$$\overline{c_j^Q} = \max\left\{\varrho_1, \varrho_2, \ldots, \varrho_j\right\} \quad \forall\, j \tag{20}$$

Equations (13) and (14), are used to maximize the network's spectral modularity values. Constraint (15) would exclude the similar sets of commenters transferred from commenter-level by $\overline{\vec{c}\delta}_{i_{n\times k}}$. Equation (16) will measure the network's spectral modularity matrix, where $d \in R^{n\times 1}$, would generate all commenters' degree values in vector representation, g is the number of commenters in the graph, and A_{ij} is the graph adjacency matrix. Equation (17) is the union between the $\vec{c}\delta_{i_{n\times k}}$ vector and the nominee sets of commenters in δ_j that the model would need to elaborate via the spectral modularity. Equation (18) is to measure the network's spectral modularity values for the network that has $\xi_j \in R^{n\times k}$ portioned and $k = \{1, 2, \ldots, n\}$ is the number of partitions that would increase iteratively.

Equation (19) enforces the spectral modularity's upper and lower boundaries (ϱ^l ,ϱ^U) respectively. Equation (20) is used to get the best joint set of commenters from δ_j that maximized the network sparsity and transfer sets to the commenter-level using parameter $\overline{c_j^Q}$.

3.5 Multi-criteria problem

Real-world small networks' criteria mentioned in Zafarani et al. (2014) were used to rank the identified sets of commenters from the bi-level analysis. This step is to verify if the sets were selected randomly or they are real-world small communities (Zafarani et al. 2014). The model will score each set of commenters with respect to five important features as presented in Eqs. (21)–(23). To originate the evaluation, each influential set was measured with respect to features such as the network clustering coefficient values, network diameter value, the network density value, the network diameter, and the network path length (Zafarani et al. 2014). For example, if a network maximized the density value, then it includes active commenters coordinating/communicating with each other, or if it minimizes the diameter then the commenters can coordinate/communicate faster (Zafarani et al. 2014).

The model utilized a multi-criteria optimization problem to measure the sets' average degree centrality values (ADC_F), average clustering coefficient values (AC), and density values (DN), where the model will assign higher rank to those sets that maximize these three criteria. In other words, the model assigns a higher weight when a set receives higher ranks or assigns a low weight when a set gets minimum rank, as presented in Eq. (22). Accordingly, the model prefers the smallest possible sets, where the multi-criteria problem is designed to measure the network's diameter and the path length respectively. The model will assign higher ranks to sets that minimize the diameter values (R_D) and the average path length (R_{Al}), as stated in Eq. (23). In other words, the model assigns higher weights to sets that minimize the diameter and path length values or assigns lower weights to sets that maximize these values.

$$\rho_{Fi} = \frac{1}{5}\left(R_{\mathcal{AC}} + R_{ADC} + R_{DN} + R_D + R_{Al}\right) \tag{21}$$

$$R_{\mathcal{AC}} = \begin{cases} W_{Fi} = 5 \;\; \mathcal{AC}_{Fi} \le x \\ W_{Fi} = 4 \;\; x < \mathcal{AC}_{Fi} \le 0.9x \\ W_{Fi} = 3 \;\; 0.9x < \mathcal{AC}_{Fi} \le 0.8x \\ W_{Fi} = 2 \;\; 0.8x < \mathcal{AC}_{Fi} \le 0.6x \\ W_{Fi} = 1 \;\; otherwise \end{cases} \tag{22}$$

$$R_{\mathcal{A}\ell} = \begin{cases} W_{Fi} = 5 & \mathcal{A}\ell_{Fi} \leq y \\ W_{Fi} = 4 & y < \mathcal{A}\ell_{Fi} \leq 1.5y \\ W_{Fi} = 3 & 1.5y < \mathcal{A}\ell_{Fi} \leq 2y \\ W_{Fi} = 2 & 2y < \mathcal{A}\ell_{Fi} \leq 3y \\ W_{Fi} = 1 & otherwise \end{cases} \tag{23}$$

Equation (21) is to measure the set's rank values ρ_{Fi}, based on the five different criteria explained earlier. Equation (22) is used to maximize (ADC_F), (AC), and (DN) respectively, where W_{Fi}, is the selected weight and assigned to set when it is close to max/min scales, and x is the actual set's scale. Equation (23) is to minimize (R_D), and (R_{Al}) respectively, where W_{Fi}, is the weight assigned to set when it is close to max/min scales and x is the actual set's scale. Also, the numeric parameters are thresholds that can be assigned by the user.

4 Toy example

Since the model contains higher level of complex operations and analysis at the commenter-level and the group-level, more clarification is needed to unblur the interactions between these two levels. In this section, a toy example is considered to observe the solution procedure steps and explain the complexity analysis.

T karate club network (Zachary 1977) displayed in Fig. 3, (includes 34 members and 78 edges), is utilized to illustrate the interactions between the commenter-level and the group-level. Where the earlier seeks to optimize the network's members' degree centrality values; and the group-level, tries to optimize the network's sparsity values. Also, this example indicates the graph's expansion, when both centrality and spectral modularity exchange information by using the assigned parameters as explained in Sect. 3.4.

Figure 7 below expresses the model's outcome after each iteration. After the model's initialization, the commenter-level will introduce the first identified central nodes from the first iteration into the group-level. This includes the most central nodes in the network such as nodes 33 and 34.

In the next iteration, the commenter-level will transfer focal structure (FSA#1) into the group-level using parameter $\vec{c}\delta_{134\times1} = FSA\#1$. Next, at the group-level the model will implement Eqs. (16) through (18) to find the best joint set of nodes that can maximize the modularity values in the network. Over and above that, the spectral modularity measures the graph's sparsity, then selects the best set of members that maximizes the modularity value ϱ_j. The results in this iteration, will identify the influential set of nodes that maximized the sparsity in the network (FSA#2). Later, the group-level will transfer this finding into the commenter-level using the parameter $c_1^Q = [1, 2, 18, 22]$. In the next iteration, the model will expand the graph from one vector (FSA#1) into the union of two sets, including the two focal structure sets (#1 and #2), and update k = 2 as presented in Figs. 7 and 8.

The same procedure will continue to discover other active sets of members and expand the graph iteratively until no new members are spared. Also, the model considered the termination conditions, when the modularity values do not change or decrease into the negative values, then the model will terminate.

Finally, the outcomes from the case study and the multi-criteria results show that (*FSA*#3) occupied the most critical position in the network and received the higher rank. The set's position will provide the set a higher influence, where nodes 1 and 4 are both central nodes, linked to higher number of members, and can control other members' influence in network.

5 Experimental results

The YouTube network described in Sect. 3.2 was employed to measure the model's performance and analyze complex real-world social networks. The model's outcomes proposed twenty-nine focal structure sets of commenters, where most of them were prominent sets hidden in the network, as shown in "Appendix". The observed outcomes suggested that these focal structure sets of commenters have the enough resources, the necessary links, and enough influence to spread fake news across the network. However, based on our text analysis performed at the end of this analysis, the model selected sets include commenters had higher level of coordination, able to maximize their influence to other users, and easily communicate with other influential commenters in different parts of the network.

For such important features, these focal structure sets of commenters would have the following characteristics:

The identified sets include active commenters acting in different focal structure sets in the network as presented in Fig. 9.
These commenters can spread false content to many other commenters and communities to coordinate the information spread across the network. Figure 9 illustrates the interactions between the commenters in each set and other influential commenters in different sets.

The achieved unsupervised model in this paper, investigated the size of the focal structure sets as part of the solution procedures', as elaborated in Sect. 6.1. In addition, Fig. 10 monitors the counts of commenters in each set, where the biggest set has one hundred and one commenters, and the smallest set includes three commenters only.

Figure 11 presents the set's average centrality values, these values monitoring how central the focal structure sets are? However, per the focal structure sets' definition in Şen et al. (2016), it is natural that some commenters may not be influential on their own, but they become influential when they coordinating with other influential commenters. Focal structure analysis is defined to find such role player commenters in small sets and count them as central commenters as other influential members inside the focal structure sets.

However, such focal structure sets of commenters have the abilities to coordinate with their active neighbors to spread false information to the maximum number of

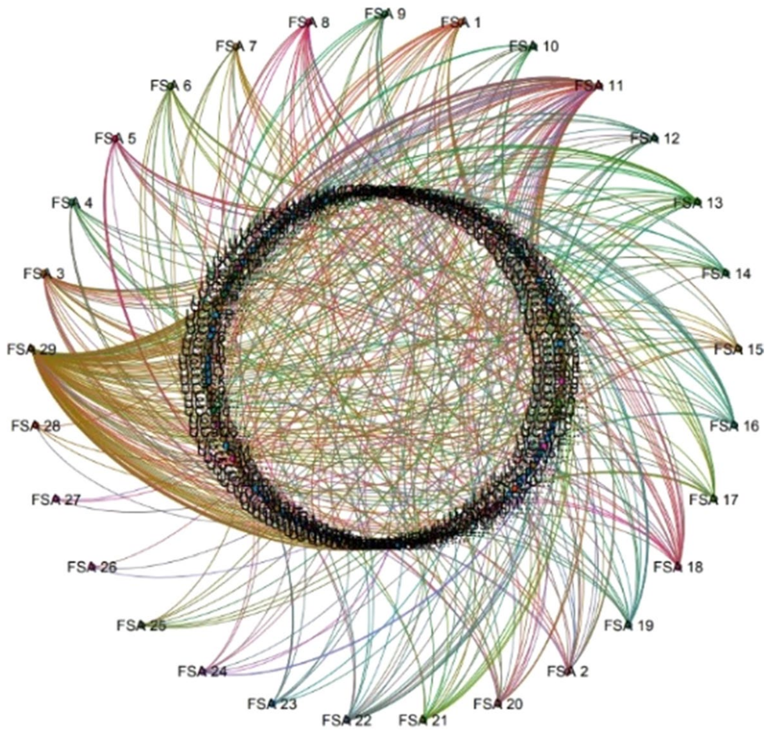


Fig. 9 Focal structure sets explored from the fake news YouTube cannel

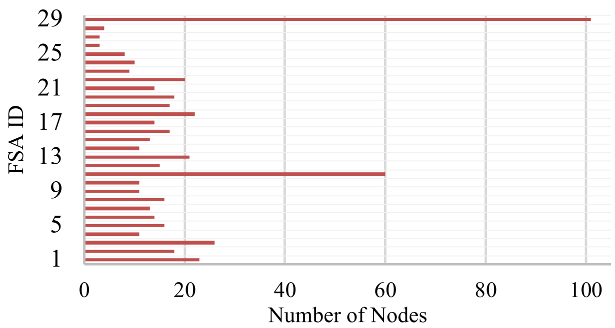


Fig. 10 Focal structure sets' size

users, post similar comments many times, reply to comments posted on the channel, and work to deviate the true facts about the posted videos.

In addition, Fig. 12 shows the how the identified sets projecting high interactions between the commenters, where the outcomes from the clustering coefficient method indicated high level of coordination between the set's commenters. These numbers will support the model's claims, where the identified sets were

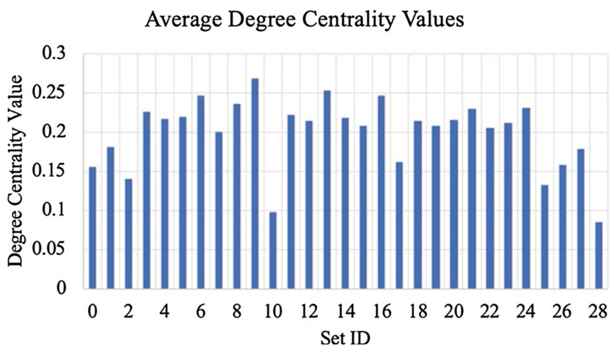


Fig. 11 Focal structure sets' average degree centrality values

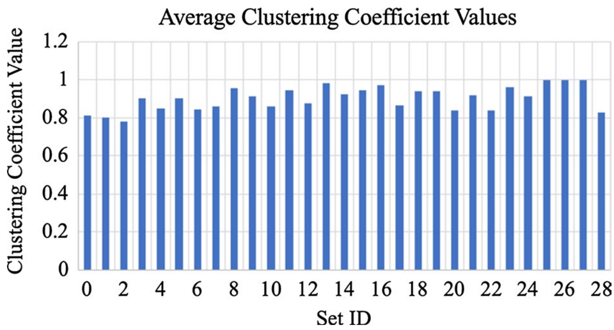


Fig. 12 Focal structure sets' average clustering coefficient values

not randomly selected by the model and include influential and central commenters coordinating with their set's members actively.

In summary, the model was able to find the focal structure sets of commenters hidden in a complex YouTube network. By the initial analysis implemented above, they are able to coordinate and spread false information about posted videos in the YouTube channel. In addition, they have the power to influence the maximum number of users in different parts of the network and impact the entire network with their agendas.

For the final analysis at this level, depending on the small real-world network criteria. Since the identified focal structure sets showed a high level of centrality and activities between the set's members, the multi-criteria problem ranked the sets based on the five different criteria explained in Sect. 3.5 as displayed in Fig. 13.

The presented model, enhanced the focal structure analysis outcomes in compare to the state of the art model in Şen et al. (2016), where the model showed promising optimized solutions and improved the focal structure analysis significantly. In addition, at the first level of analysis, the model definitely optimized the commenter-level to uncover the influential commenters connected to active

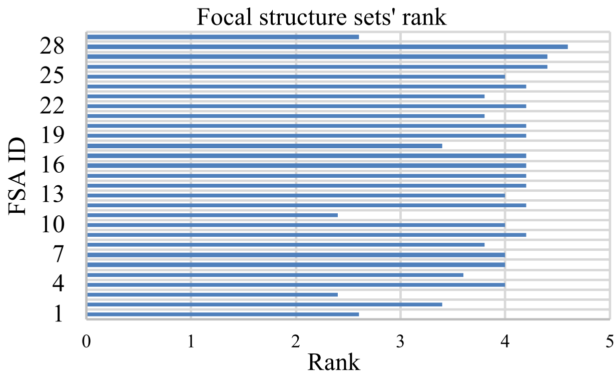


Fig. 13 Focal structure sets have high ranks

neighbors. Then in the second level of the analysis, based on previous level's outcomes, the model identified the active groups that maximized the sparsity in the networks. Next section is to present the validation methods.

6 Validation

Since the model's outcomes demonstrated a significance focal structure analysis, then this study will contribute the initial and practical step towards blocking the fake news spread in complex social networks. To begin such analysis, we have utilized two methods for measuring the sets' impact, commenters' behaviors, and the commenters' coordinated actions in both local and global aspects.

The results in this section would help to validating the findings and support the model's developments by evaluating the sets' performances at different levels. Moreover, this model contributes to the efforts where makes is easier to suspend malicious sets of commenters than random central users. This will definitely advantage the model to stop fake news hubs and identify the information disseminators in social networks. Likewise, the model's results can offer information about the set's commenters' location(s), what parts of the social network are the targets to conspiracy theory? and observe the strongest and weakest locations for information dissemination in the network. The results can help any network's analysts to measure the influence generated by each focal structure set and determine which sets should be suspended to stem the disinformation spread.

6.1 Network sparsity

This section is implemented to measuring the impacts generated by the focal structure sets in social networks. We determine if the proposed model can truly identify the fake news spreaders and helps to block the information dissemination caused by those focal structure sets. Also, the endeavor to test the model's applicability towards the initial goal and to suspend the most influential groups, preventing them from spreading false information and enflaming other radical actions within the social networks.

For this purpose, Girvan–Newman modularity method (2002b) is utilized to measure the changes when any influential set is suspended from the YouTube network. This step helps to expose more information about each focal structure set; such as the sets' structure in the network, links and resources, commenters' activities, number of posted comments, number of followers, and the number of communities any focal structure set can access.

Initially, Girvan–Newman method (2002b) clustered the network into 160 communities before suspending any focal structure set. Then, the model suspended one focal structure set at a time to observe the changes in the network. However, the point here is about the sets' structure in the network, if blocking any focal set will dramatically increase the number of communities and divides the network into a larger number of communities, then this is an influential focal structure, and it is able to spread information across the network easily. In addition, the focal structure set's size does matter here as illustrated in Fig. 10, where if a small set can make a remarkable increase in the number of communities, then it indicates that this set has more influence than bigger focal sets have lower impacts on the network's sparsity.

Table 1 displays the changes in the network after blocking each focal structure set, where the number of communities increases in the interval of [279–818] communities. Also, Table 1 proves that a small focal structure set such as (*FSA*#27) consisted of only three commenters was able to influence higher number of users and have more influence than other larger focal structure sets as shown in the table below.

6.2 Weakly connected commenters

In this part, we measure the local impacts generated by each focal structure set to cover questions like: What can a focal structure set can do to other users' connectivity? Can a focal structure set connect commenters from different parts of the network? Can a focal structure set control the information flow across the network? And can a focal structure set (dis)connect users (from) in the network? For this purpose, we utilized the depth first search method proposed by Tarjan et al. (1972). This method will help to measure the weakly connected commenters after suspending the focal structure sets from the YouTube network.

The initial outcome displays 142 weakly connected commenters before suspending any focal structure set from the network. However, after suspending the focal structure sets, the results showed a huge increase in the numbers of weakly connected commenters across the network.

Table 1 signifies the changes in the communities' structures when each focal structure set was suspended from the network. For example, focal structure set (*FSA*#1) was able to disconnect more than 680 users from the network as presented in Table 1. This fact makes it clear that these focal sets include commenters having strong communications with the influential commenters in the given social network. Another example to mention here is (*FSA*#11), it was able to

Table 1 Focal structure sets' impact within the network

FSA ID	Communities	Weakly connected commenters
FSA 1	699	682
FSA 2	650	631
FSA 3	663	646
FSA 4	276	260
FSA 5	629	615
FSA 6	279	264
FSA 7	577	560
FSA 8	304	280
FSA 9	291	275
FSA 10	576	558
FSA 11	818	802
FSA 12	324	308
FSA 13	647	631
FSA 14	274	259
FSA 15	289	272
FSA 16	301	286
FSA 17	303	286
FSA 18	317	303
FSA 19	321	302
FSA 20	305	289
FSA 21	581	559
FSA 22	645	628
FSA 23	256	237
FSA 24	262	244
FSA 25	252	233
FSA 26	170	152
FSA 27	477	461
FSA 28	175	159
FSA 29	673	653

influence more than 802 users. In addition, blocking this focal structure set (60 users) from the network is way more efficient than suspending random central commenters.

It is clear that such small focal sets can act faster and more effective on the conspiracy theories and suspending them will help to stop the spread to hundreds and thousands of users. Blocking the focal structure sets will help to sparse the network into hundreds of smoother communities that makes it easy to analysis the groups' movements.

6.3 Focal structure sets' behavior

This section will analyze the focal structure sets' behaviors and classify the sets' discussion based on the commenters' actions. In other words, this section is to verify if the identified focal structure sets actually spread fake news about the conflict in Soh China Sea, if the focal structure sets injected radical comments to influence others with fake facts or the other way, and if they posted peaceful or true fact comments. Figure 14 shows the statistics for each focal structure set extracted from the dataset. This figure is demonstrating the sets' size, number of comments, and number of comments per video per set in the YouTube Channel. Next, we charted the common topics/words used by all focal structure sets' commenters; words/topics such as "South", "China", "Chinas", "Chinese", "Sea", "UNCLOS" (The United Nations Convention on the Law of the Sea), "SCS" (South China Sea), "JAPAN", "Taiwan", "USA", and "war" were mentioned the most.

In addition, the interesting point from the outcomes is that the model clustered the influential commenters with similar interests into different focal structure sets, where these influential commenters were coordinating to spread parallel and related information across the network. For example, (*FSA*#27), includes of only three commenters (UC80xWEM3NOB_KLLHQPCmEyw, UC2zEIGq0XpGUv_cnZJAljeA, and UC2Qo_PtpbJNep-AViyFCH2g), posted comments related to topics like war, radical actions, and fake news about the policies in China, the United States, and other countries. In addition, all the commenters did comment on topics related to the military, President Trump, trade war, ships, jets, war started in the region, and other radical topics as presented in Table 2. Correspondingly, all members in this set supplemented each other's comments through posting fake news and comments about war is happening in the coming days, how the Chinese army and the United States navy are involved in this conflict, and the war will start by bombing the important places in the region.

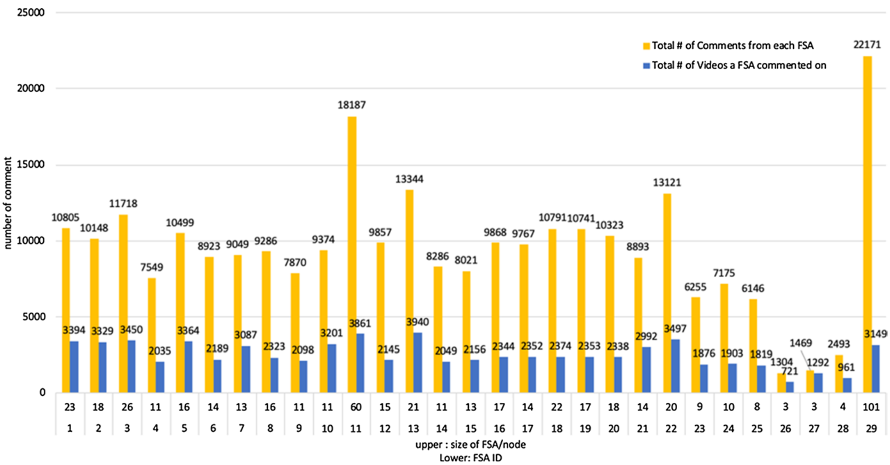


Fig. 14 Focal structure sets statistics extracted form the dataset

Table 2 Sample of comments posted by focal structure sets' commenters (FSA # 27), where *UC80x-WEM3NOB_KLLHQPCmEyw* posted about the war in the region, types of weapons, and involving other countries in this conflict

Commenter ID	Comments samples
UC80xWEM3NOB_KLLHQPCmEyw	*"The aircraft is powered by four WJ6 turbofan engines. In addition, there are sources that China may use the previously purchased D-30KP-2 engine in large numbers in Russia, for installation on H6 bombers, transport aircraft. Russia Il-76 and Y-20 domestic transport aircraft" *"China is looking to promote trust between the two countries" *"Mr. Trump described China as enemies of the United States, and pledged against Beijing"
UC2zEIGq0XpGUv_cnZJAljeA	*"Chinas policy toward South China Sea since Deng Xiaoping is lay the disputes aside and exploit together. China has a long historical relationship with S China" *"Gary I am an average Joe! Nothing related with CCP or government. But I feel necessary to show the facts concerning distortion of fact about China!"
UC2Qo_PtpbJNep-AViyFCH2g	*"Philippines president shouldnt allowed china is terrorist comminest dont trusted one you country will invation by Chinese comminest dictorship greedy for land" *"BrotherMohd Hanif another war starting china will desroy isreal and usa will congereing the world india japan France European countries will joins with USA bring to ISREALS USA control"

The other two commenters supplemented him by comments about China's actions and the conflicts between China and the United States

The results of this analysis, the frequently posted comments by other focal structure sets demonstrated the behaviors and the levels of the discussions in this channel. However, after an in-depth analysis and depending on the discussions' levels, we categorized the identified focal structure sets into three categories as follows:

- The first category includes the offensive sets of commenters who frequently posted about wars, the connection to the United States, how oil/natural resources are in danger. We saw commenters spreading fears to all citizens and how they will go under the USA navy's attacks. These sets were spreading fake information about weapons and other aggressive topics.
- The second category includes focal structure sets posting about other countries' presidents and political figures. In this level of discussion, major number of commenters were debating on how other countries influencing the conflict. We found sets continuously pointing to President Trump, Philippine's Prime Minister Lee, and Philippine's President Duterte. Likewise, comments about countries such as Taiwan, Japan, China, and the Philippines were involved in this level. Agencies like UN and other active national associations in the region such as EEZ (An Exclusive Economic Zone), ASEAN (Association of Southeast Asian Nations), etc. were blamed by not taking any serious actions

to end the conflict. However, such sets were trying to push for radical behaviors against other countries and other organizations.

- The Third category includes sets were spreading info in a completely different direction. These commenters were posting about peace and steps that could help to avoid any comping war. Also, these focal structure sets posted comments related to the positive aspects of negotiations, involving and obey the international courts, praying to God to prevent any war in the region, and reminding others about all the suffers if the war starts.

7 Conclusion and discussion

In this paper, we proposed the extended focal structure analysis model to identify hidden influential sets of commenters in a complex YouTube channel. The model optimized (*max–max*) the user-level community detection algorithm (degree centrality) and the group-level detection algorithm (spectral modularity) respectively. In addition, the model utilized other Graph theory measures to improve the performance and rank the extracted groups for higher quality.

Two methods were used to validate the results at both local and network levels respectively. In addition, we utilized the text analysis methods to study the sets' behaviors, commenters actions, and the overall discussion's directions to provide larger pictures of the conspiracy theories established by the identified focal structure sets. This work is an extension to our prior research in Alassad et al. (2019b) where in this work we utilized new methods to validate the results, measuring network's resiliency, complexity analysis, and how any focal structure sets' commenters were behaving in the network.

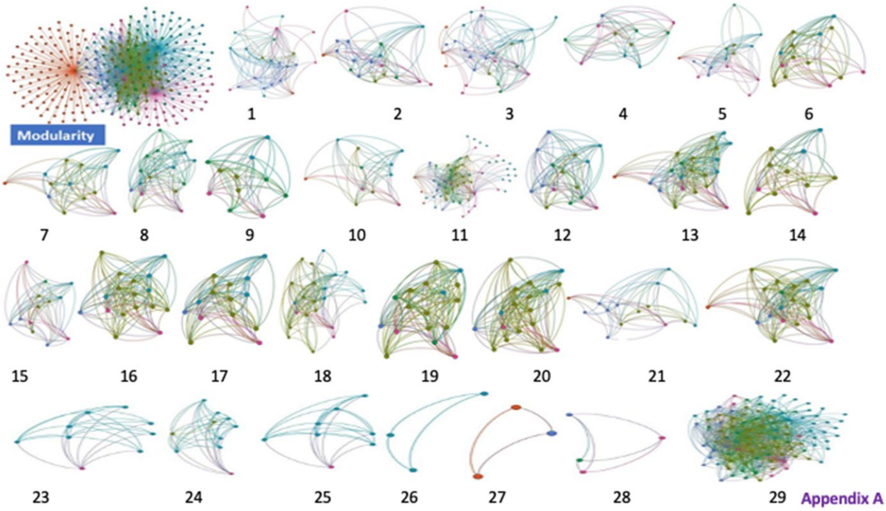
Likewise, the model proposed the active sets of commenters as key connectors that can spread information to the maximum number of users and able to (dis)connect the maximum number of users to (from) the network. Moreover, this model is an unsupervised model and does not require any information/parameters from the users as inputs into the analysis (e.g., number of clusters, lower and upper bounds), except for the unimodular network dataset (source, target). In other words, it is a credit to propose such unsupervised model able to provide an intensive analysis with limited amount of information. Also, the model was able to overcome the drawbacks mentioned in the state of the art model presented by Şen et al. (2016). The model proposed in this research was skilled to apply and illustrate the definition of the Focal Structure Analysis stated in Şen et al. (2016). This model successfully identified the smallest possible key intensive communities including members that span multiple focal structure sets, and acting in different parts of the social networks. In addition, this analysis demonstrated the model's efficacy by providing a practical method to stop information dissemination in complex social networks.

Moreover, this study highlights interesting contributions into the social network analysis listed as follows:

- This research sheds new light on key sets of commenters commenting to escalate the tension in the social network such as the YouTube channel spreading fake news about the South China Sea conflict.
- This study reveals, and clusters hidden key sets of commenters collaborating for posting similar comments about the war, violence, and radical behaviors to provoke online hysteria.
- The proposed model identifies key sets of commenters that posted false comments about other countries' policies and actions that they would take to participate in any upcoming war, possible bombing, and unusual economic sanctions.
- The model clusters key sets of commenters spreading false information about decisions related to other countries sending more troops, more jets, or submarines in conflict areas.
- Finally, this research labels key sets of commenters spreading peaceful comments, trying to bring wisdom to the table through the UN, involving the courts, and other professional national associations and law firms. Moreover, these groups spread comments related to the consequences of any upcoming war in the region along with anti-fake news comments and real facts to debunk the fake news.

The practical side in the identification of influential users and their focal structure it to allow us to suspend only the accounts within social media networks that have the most structural influence for information flow instead of selecting random users or only the most central users in the network. However, the main limitations of this research are related to the users' connectivity, the network's structures, and if the adversaries change their strategies in the network. For this purpose, the next steps in the research are (1) optimize the solutions between different available centrality methods (e.g., betweenness centrality), (2) investigate a dynamic version of focal structure analysis to proactively track changes in adversaries' strategies.

Appendix



.

Acknowledgements This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (Award Number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

Alassad, M., Agarwal, N., Hussain, M.N.: Examining intensive groups in YouTube commenter networks. In: Proc. 12th Int. Conf. SBP-BRiMS 2019, no. 12, pp. 224–233 (2019a)

Alassad, M., Hussain, M.N., Agarwal, N.: Finding fake news key spreaders in complex social networks by using bi-level decomposition optimization method. In: International Conference on Modelling and Simulation of Social-Behavioural Phenomena in Creative Societies, pp. 41–54 (2019b)

Alassad, M., Hussain, M.N., Agarwal, N.: Developing graph theoretic techniques to identify amplification and coordination activities of influential sets of users. In: Accept. Int. Conf. Soc. Comput. Behav. Model. Predict. Behav. Represent. Model. Simul., (2020)

Alinezhad E, Teimourpour B, Sepehri MM, Kargari M (2020) Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches. Neural Comput. Appl. 32(8):3203–3220

Briscoe, E.J., Appling, D.S., Mappus, R.L., Hayes, H.: Determining credibility from social network structure. In: Proc. 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min., pp. 1418–1424 (2014)
Chan, E.Y.K., Yeung, D.Y.: A convex formulation of modularity maximization for community detection. In: IJCAI Int. Jt. Conf. Artif. Intell., pp. 2218–2225 (2011)
Chen, W., Wang, Y.: Efficient influence maximization in social networks categories and subject descriptors. In: Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 199–207 (2009)
Chen N, Liu Y, Chen H, Cheng J (2017) Detecting communities in social networks using label propagation with information entropy. Phys. A Stat. Mech. Appl. 471:788–798
Dinh, T.N., Li, X., Thai, M.T.: Network clustering via maximizing modularity: approximation algorithms and theoretical limits. In: Proc. IEEE Int. Conf. Data Mining, ICDM, vol. 2016-Janua, pp. 101–110 (2016)
Girvan M, Newman MEJ (2002a) Community structure in social and biological networks. Proc. Natl. Acad. Sci. 99(12):7821–7826
Girvan M, Newman M (2002b) Community structure in social and biological networks. PNAS 99(12):7821–7826
Herzig, J., Mass, Y., Roitman, H.: An author-reader influence model for detecting topic-based influencers in social media. In: Proc. 25th ACM Conf. Hypertext Soc. media, pp. 46–55 (2014)
Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: Elev. Int. AAAI Conf. Web Soc. Media, pp. 759–766 (2017)
Hu F, Liu Y (2016) A new algorithm CNM-Centrality of detecting communities based on node centrality. Phys. A Stat. Mech. Appl. 446:138–151
Hussain, M.N., Tokdemir, S., Agarwal, N., Al-Khateeb, S.: Analyzing disinformation and crowd manipulation tactics on YouTube. In: 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min., pp. 1092–1095 (2018)
Izunaga Y, Yamamoto Y (2017) A cutting plane algorithm for modularity maximization problem. J. Oper. Res. Soc. Jpn. 60(1):24–42
Jones, S., O'Neill, E.: Feasibility of structural network clustering for group-based privacy control in social networks. In: Proc. Sixth Symp. Usable Priv. Secur., p. 9 (2010)
Kleinberg JONM (1999) Authoritative sources in a hyperlinked environment. Proc. ACM-SIAM Symp. Discrete Algorithms 46(5):604–632
Leskovec, J., Mcglohon, M., Faloutsos, C., Glance, N., Hurst, M.: Patterns of cascading behavior in large blog graphs. In: Proc. 2007 SIAM Int. Conf. data Min., pp. 551–556 (2007a)
Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 551–556 (2007b)
Li C, Wang L, Sun S, Xia C (2018) Identification of influential spreaders based on classified neighbors in real-world complex networks. Appl. Math. Comput. 320(11):512–523
Newman MEJ (2004a) Fast algorithm for detecting community structure in networks. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. 69(6):5
Newman MEJ (2004b) Detecting community structure in networks. Eur. Phys. J. B Condens. Matter 38(2):321–330
Newman MEJ (2006) Modularity and community structure in networks. Proc. Natl. Acad. Sci. 103(23):8577–8582
Page L, Brin S, Motwani R, Winograd T (1998) The PageRank citation ranking: bringing order to the Web. World Wide Web Internet Web Inf. Syst. 54(1999–66):1–17
Sato, K., Izunaga, Y.: A branch-and-price approach with MILP formulation to modularity density maximization on graphs, pp. 1–25 (2017). arXiv Prepr. http://arxiv.org/abs/1705.02961
Şen F, Wigand R, Agarwal N, Tokdemir S, Kasprzyk R (2016) Focal structures analysis: identifying influential sets of individuals in a social network. Soc. Netw. Anal. Min. 6(1):17
Tarjan R (1972) Depth-first search and linear graph algorithms. SIAM J. Comput. 1(2):146–160
Tsung, C.K., Ho, H., Chou, S., Lin, J., Lee, S.: A spectral clustering approach based on modularity maximization for community detection problem. In: Proc. 2016 Int. Comput. Symp. ICS 2016, pp. 12–17 (2017)
Waltman L, Van Eck NJ (2013) A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B 86(11):1–14

Xie WB, Lee YL, Wang C, Chen DB, Zhou T (2020) Hierarchical clustering supported by reciprocal nearest neighbors. Inf. Sci. (Ny) 527:279–292
Yang L, Silva JC, Papageorgiou LG, Tsoka S (2016) Community structure detection for directed networks through modularity optimisation. Algorithms 9(4):1–10
You X, Ma Y, Liu Z (2020) A three-stage algorithm on community detection in social networks. Knowl. Based Syst. 187:104822
Zachary WW (1977) An information flow model for conflict and fission in small groups. J. Anthropol. Res. 33(4):452–473
Zafarani R, Abbasi MA, Liu H (2014) Social Media Mining: An Introduction. Cambridge University Press, Cambridge

Authors and Affiliations

Mustafa Alassad[1] · Muhammad Nihal Hussain[1] · Nitin Agarwal[1]

✉ Nitin Agarwal
nxagarwal@ualr.edu

Mustafa Alassad
mmalassad@ualr.edu

Muhammad Nihal Hussain
mnhussain@ualr.edu

1 University of Arkansas at Little Rock, Little Rock, AR, USA