Social Network Analysis and Mining (2022) 12:103
https://doi.org/10.1007/s13278-022-00938-0

ORIGINAL ARTICLE

Contextualizing focal structure analysis in social networks

Mustafa Alassad[1] · **Nitin Agarwal**[1]

Received: 28 February 2022 / Revised: 28 June 2022 / Accepted: 15 July 2022

Abstract
Focal structures are key sets of individuals who may be responsible for coordinating events, protests, or leading citizen engagement efforts on social media networks. Discovering focal structures that are able to promote online social campaigns is important but complex. Unlike influential individuals, focal structures can affect large-scale complex social processes. In our prior work, we applied a greedy algorithm and bi-level decomposition optimization solution to identify focal structures in social media networks. However, the outcomes lacked a contextual representation of the focal structures that affected interpretability. In this research, we present a novel contextual focal structure analysis (CFSA) model to enhance the discovery and the interpretability of the focal structures to provide the context in terms of the content shared by the focal structures through their communication network. The model utilizes multiplex networks, where one layer is the user network based on mentions, replies, friends, and followers, and the second layer is the hashtag co-occurrence network. The two layers have interconnections based on the user hashtag relations. The model's performance was evaluated on various real-world datasets from Twitter related to COVID-19, the Trump vaccine hashtag, and the Black Lives Matter (BLM) social movement during the 2020–2021 time. The model discovered contextual focal structures (CFS) sets revealed the context regarding individuals' interests. We then evaluated the model's efficacy using various network structural measures such as the modularity method, network stability, and average clustering coefficient to measure the influence of the CFS sets in the network. Ranking correlation coefficient (RCC) was used to conduct the comparative evaluation with real-world scenarios to find the correlated solutions.

Keywords Multiplex Networks · Complex Network · Focal Structures · Entropy · Information Gain · COVID-19 · Contextual Focal Structures

1 Introduction

Social media platforms like Twitter, Instagram, and Facebook are fast-growing microblogging services that report daily news, social activities, and local/global real-life events. In addition, these platforms became very popular, where Twitter and other platforms let their users exchange information, links, images, or videos with limited and no restrictions on the contents. For example, during recent health crises (COVID-19 pandemic) and other social events like (the US 2020 election), millions of users on Twitter reported their experiences, shared their thoughts on fighting COVID-19, (dis)agreed with COVID-19 regulations, and predicted the results of the US 2020 election. Furthermore, due to ease of use, Twitter turns out to be one of the favorite online platforms and communication channels, where the coordinating groups spreading conspiracy theories related to events like the Black Lives Matter (BLM) movement, Make America Great Again (MAGA), and COVID-19 anti/vaccination plans. Therefore, online social media platforms are converted into important sources for true or false information, fake news, misinformation, and disinformation that could benefit or damage society at the same time.

On the other hand, there is a need to discover such coordinating groups and filter out the shared information. Many different online social network analysis methods were used to reveal the meaning of the disseminated information and simplify the connections between users or communities at a micro-level. Therefore, robust analysis is required to enable

✉ Mustafa Alassad
mmalassad@ualr.edu

Nitin Agarwal
nxagarwal@ualr.edu

[1] COSMOS Research Center, UA - Little Rock, Little Rock, AR, USA

Published online: 08 August 2022

the individuals, agencies, and government organizations to stay informed of "what is happening between online users/communities?" However, due to the high complexities in the structure of online social networks, it is challenging to analyze every single movement of millions of users and track the evolution of the dynamic communities over time. In this matter, one interesting topic that needs detailed analysis is the discovery of the online coordinating groups spreading information to influence the maximum number of users on social networks. These coordinating groups include users acting in different communities, and when they operate together, they can maximize their influence, mobilize crowds, and organize online campaigns (Şen et al. Dec. 2016; Alassad et al. 2021a). Thus, such online coordinating groups have unique structures in the body of the social network and could occupy central users to maximize the influence and the information spread on social networks.

Therefore, identifying the coordinating groups on social networks is the main contribution to this research, where these groups could develop unique structures and act to influence individuals/communities to maximize the information dissemination across social networks. Nevertheless, conventional community detection methods focus on larger communities and are oblivious to these coordinating groups. Moreover, as the social networks exponentially grow, their structure becomes complex, and communities reorganize, making it challenging to identify these coordinating groups and their information diffusion networks (Şen et al. Dec. 2016; Alassad et al. 2019a).

To fill the gap in the analysis, Şen et al. (Şen et al. Dec. 2016) proposed the focal structure analysis (FSA) model to identify the smallest possible influential (coordinating) groups of users that can maximize the information diffusion in social networks. However, this model suffered from a few drawbacks, such as low-quality focal structure sets (chain groups), limited users' connections to only one set, and no information was provided about the FSA sets' contextual activities. In similar research, Alassad et al. (Alassad et al. 2019a) introduced the FSA 2.0 model to enhance the quality of the focal structure sets discovery and to overcome the limits in the activities of the influential users. For this purpose, the authors developed a bi-level decomposition optimization model to identify groups that could maximize the individual's influence in the first level and measures the network's stability in the second level. Nevertheless, the FSA 2.0 model presented in Alassad et al. 2019a used only a unimodular user–user network in the analysis, where the outcomes were missing the context activities, the users' interests, and overall behavior of the focal structure sets as explained in this paper. To overcome these drawbacks in the state-of-the-art model mentioned in Şen et al. Dec. (2016) and FSA 2.0 in Alassad et al. 2019a, the contextual focal structure analysis model (CFSA) should reveal the context activities of the focal structure sets and highlights the behavior and users' interests on social networks. To simplify the introduction to the CFSA model, we implemented an illustrative example to show the development of the model and its advantages of the CFSA model.

1.1 Illustrative example

There is a need to identify online coordinating groups, understand users' interests, study the interactions of individuals, and determine focal structures' influence on social networks. In this matter, the CFSA model shall enhance the discovery and interpretation of the focal structures (coordinating groups) to present the contexts of specific interaction patterns or communication structures on social networks.

This section demonstrates the development of the focal structure analysis models over the last few years and the development from FSA 2.0 model to the CFSA model in social network analysis. Fig. 1 shows a dataset retrieved from a large Twitter dataset of users with 94 domestic terrorism accounts sharing hashtags like "MAGA," "antifa," "trump2020," "election2020," "maga2020," and "blm,". This multiplex network consists of 74,764 vertices and 94,706 edges. In this case study, we implemented this dataset, compared both models' outcomes, and showed the enhancement in the results.

Focal structure analysis model (FSA 2.0) this model implements a unimodular network of users–users connections (6910 nodes, 12,957 edges) to identify FSA sets in the social network. In this domain, we successfully identified key groups of users that influence the maximum number of users in the network. The model combined two well-known social network analysis methods, namely centrality and modularity, to bridge the shortcomings of traditional community detection methods used in graph theory. The resultant combination is a bi-level linear optimization problem to realize/observe user-level and network-level interactions, as shown in Alassad et al. 2021b.

In addition, the FSA 2.0 model identified the most minor possible coordinating groups in this dataset, where for easiness, we show only three sets in Fig. 1 (FSA1, FSA2, and FSA3). To elaborate more, these FSA sets include users acting in different parts of the network. When working together, they can maximize the information dissemination to most users in the network. For example, FSA1 shows a link between "realDonaldTrump" and "JordanRoberts" or "ballalert" and other users; however, we only see connections between users based on mentions, friendships, and likes in the network without additional related information about the context or other activities in the network. Furthermore, another limitation is connected to the users' having common interests in the network. The FSA model does not show links between users posting or sharing similar topics or common

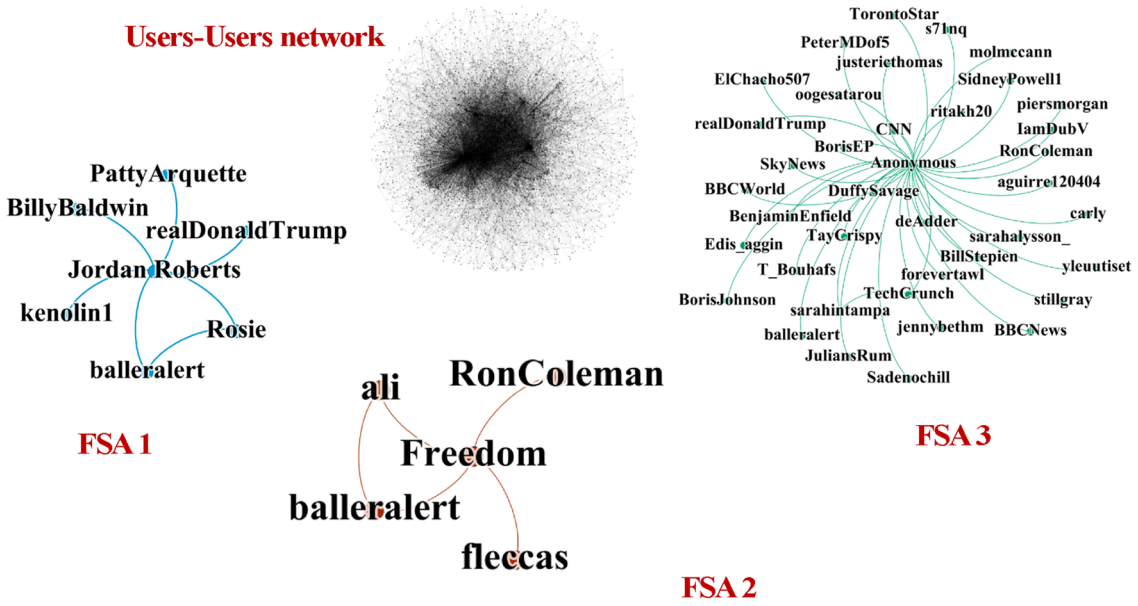


Fig. 1 FSA model outcomes

interests in the network, but the CFSA model should handle it easily. Likewise, the same limitations applied to other FSA sets are shown in Fig. 1.

Contextual focal structure analysis model (CFSA) to expand the analysis and reveal other behavior of the FSA sets, the model proposed here implements the multiplex network approach to consider other behavior of the users in the solution procedure. The CFSA model utilizes the participation layers to observe the users' behavior and interests in the network. For this purpose, the multiplex networks that include the users' activities as hashtags would reveal the interests and information shared between users on social networks. For example, Fig. 2 shows three contextual focal structure sets (CFS1, CFS2, CFS3) identified from the users–hashtag layer (participation layer) that involve users acting in different parts of the network, users sharing similar hashtags, and more reveal information about the communities' activities on the social network. The users–hashtags layer (participation layer) is the union of the users–users co-occurrence and the shared hashtags in the Twitter dataset, as explained in Sect. 4. For example, CFS1 shows context (hashtags) shared about Bill Gates, where users like "mommaesq" and "BWaveResist2020" both have links to the "realDonaldTrump" account, wherever these two users shared hateful hashtags on Bill Gates and Dr. Antony Fauci on Twitter as presented in Fig. 2. Likewise, CFS2 and CFS3 are influential sets that show users' connections, as well as the content shared on the Twitter network, as described in this paper.

In summary, FSA 2.0 model utilizes only one layer (users–users) and does not provide or observe extra information about the users' activities, shared topics, or context distributed by FSA sets on social networks. On the other hand, the CFSA model would help better interpret the focal structure sets' activities by including the participation layer that incorporates the contextual information regarding the users' interests and the focal structure sets' activities. This research consists of hashtags shared by online users (participation layer) or (Hashtags–Hashtags layer) and provides a better solution than just the users–users connections. Likewise, the CFSA model will find out the hashtags linked to users and hashtags shared by other hashtags and come up with a similar representation that would enhance the analysis on social networks.

Moreover, the CFSA model is superior to other models for a few reasons. First, this model leverages context to improve the discovery and interpretability of focal structure sets in social networks. The second reason is to use multiplex networks formalism to implement the idea, where layers could be used to represent different contexts such as relations among individuals on other platforms (Twitter, blogs, Facebook, and YouTube). Similarly, the CFSA model could implement the content/information flow network (e.g., reply/mention/share network, hashtag network, topic network, link/URL graph, and metadata inferred covert connections among entities). The layers in the CFSA model could have interconnections based on user–user and user–content relations on social networks. The third reason is to discover the unfriended users who share similar thoughts on social networks; for example, CFS3 shows accounts like "Black Lives Matter" and "Doll.face" have no user–user connections, but both accounts were interested in

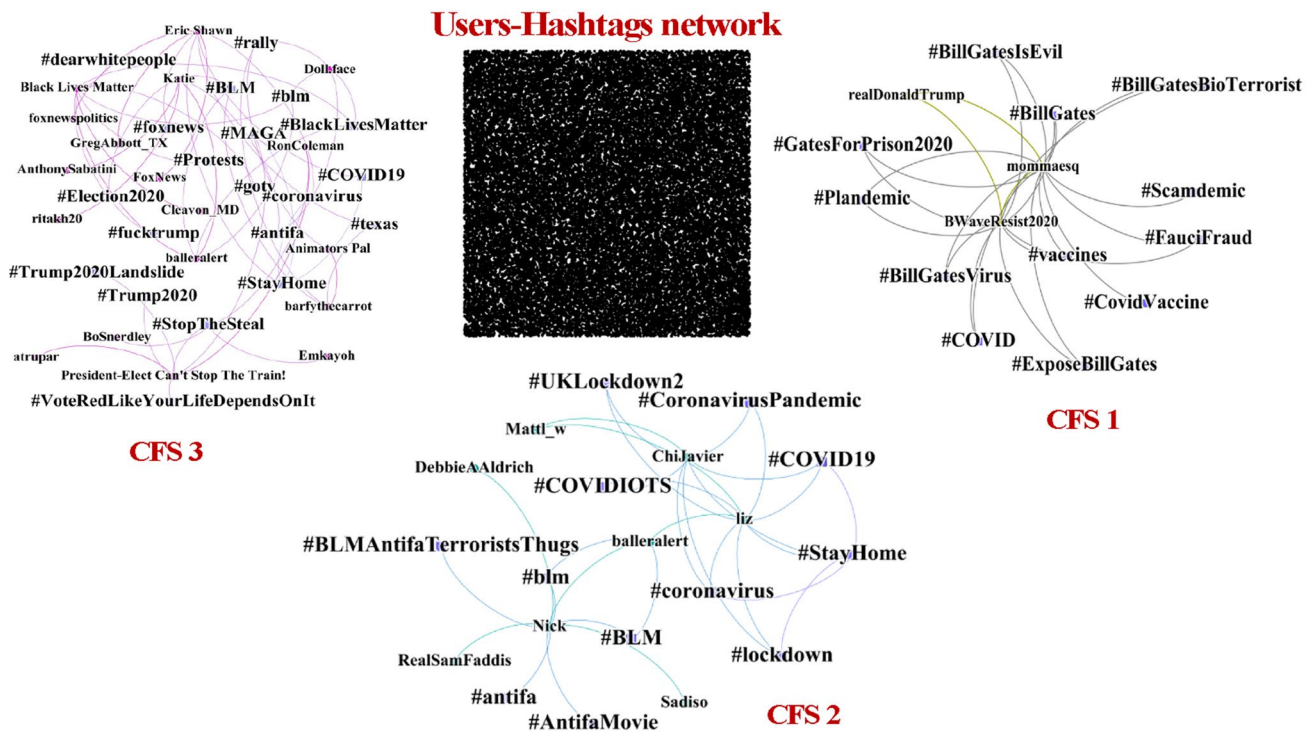


Fig. 2 Outcomes of the CFSA model

sharing "#BlackLivesMatter" hashtag. The fourth reason is finding communities with different opinions, observing the users' interactions, and context exchange on social networks.

1.2 Outline

The rest of the paper is organized as follows: Sect. 2 presents the related work on identifying focal structures and context analysis in social networks. Sect. 3 defines the problem statement and CFSA model in detail. Sect. 4 uses two real-world Twitter datasets to evaluate the model's performance. The validation procedure confirms the enhancement in the focal structure analysis modeling and real-world datasets; we concluded that the contextual focal structures analysis model (CFSA) outcomes are more interpretable and informative than the FSA 2.0 model. Sect. 5 reviews the main findings and theoretical and practical implications. Lastly, the conclusion, limitations, and directions for future research are presented in Sect. 6.

2 Related work

Many recent interesting community detection algorithms have been proposed to overcome gaps in traditional social networks analysis and handle the complexities in the behaviors of online users, as reviewed in this section.

2.1 Identifying focal structure sets in social networks

Identifying communities on social networks has become essential in many online events like anti-government movements, political and election campaigns, misinformation, disinformation, and conspiracy theories dissemination (Al-Khateeb and Agarwal 2014). Likewise, online community detection on social networks has gained close attention because of widespread online users using social media platforms ("Demographics of Social Media Users and Adoption in the United States | Pew Research Center" 2021). In the literature, we can find several methods designed to detect communities, cluster complex network structures, and optimize the patterns of well-connected users to simplify the analysis. Starting with local centrality methods see communities based on an initial division method and then the modularity method to get the final partitions in the complex social networks (Li et al. Oct. 2019). Similarly, the inverse modeling based on a multi-objective algorithm (Zou et al. Jan. 2019), local search strategy (Moradi and Parsa Jun. 2019), a multi-agent genetic algorithm (Li and Liu May 2016), the optimum ratio for clustering complex networks (Hagen and Kahng 1992), and label propagation and fuzzy C-means techniques were used to simplify the communities' discovery in complex social networks (Chen et al. Apr. 2017). All the studies mentioned above represent robust strategies for dealing with

different community detection and enhancing the complex social network analysis.

This research aims to identify influential sets of coordinating groups of users spreading information on social media platforms. Some of these sets could disseminate harmful information and cause damage to society under varying circumstances on social media platforms. Şen et al. introduced focal structure sets on social media; according to the state-of-the-art model, a focal structure set is defined as a "key set of individuals who may be responsible for organizing events, protests, or leading citizen engagement efforts" (Şen et al. Dec. 2016). The authors applied a greedy algorithm to discover the smallest possible sets involving users responsible for influencing thousands on social media platforms like Twitter and Facebook. In the same subject, a bi-level centrality-modularity model was applied to examine intensive groups of co-commenters spreading fake news on a YouTube channel (Alassad et al. 2019a). In this research, the model maximized the local centrality of the users and modularity values of the entire network to enhance the discovery of focal structure sets in social networks. Likewise, a fake YouTube news channel discovered key sets of coordinating information spreaders on social media platforms (Alassad et al. 2019b). The model used a decomposition optimization method to find focal structure sets disseminating harmful information about South China Sea conflict. Furthermore, focal structure sets, including aggressors coordinating cyber threats to intelligent infrastructure networks discovered in Alassad et al. 2021a; the authors used computational social science and the deviant cyber flash mob detection method to find sets of aggressors and measured their power to influence others on social media platforms.

However, the outcomes from these methods provide limited analysis, where these studies considered only the users–users interaction from a unimodular network in the solution procedure. Yet, to better understand the focal structures' activities, these methods are short on providing any contextual activities performed by the focal structure sets. To overcome this gap in the analysis, we introduced the CFSA model, which leverages hashtags as a context to improve the discovery and interpretability of focal structures in social networks. We implement the idea of the multiplex network and the decomposition optimization problems that could enhance the focal structure sets analysis. This model combines several users' behavior on social media platforms; considering one behavior (layer) could be the users–users network based on mentions, replies, friends, the followers, and the second action (layer) could be the hashtag co-occurrence network.

2.2 Multiplex networks

Multiplex networks are more informative in real-world applications (Luo et al. 2020) and are used to capture high levels of complexities in the relations and interactions between communities on social networks (Falih and Kanawati 2015). Many algorithms in the literature use multiplex networks to detect communities; locally adaptive random transition algorithms are based on a random walk between different layers, as mentioned in Magnani et al. Jun. (2021). Infomap is used to analyze the information flow as a random walk with teleportation in the multiplex network (Bothorel et al. Sep. 2015), random walk in multiplex networks to find relevant communities in all layers (Magnani et al. Jun. 2021), and cross-layer edge clustering coefficient approach to split communities on social networks (Guimerà et al. 2007). In addition, the cross-layer method helps clarify different behaviors, simplify the information diffusion analysis and speed up or slow down the information spread in multiplex networks (Li et al. 2015). Similarly, the attributed multiplex graph model was designed to model different relations and objects on users and edges in social networks (Hu, et al. 2020).

Moreover, each layer in multiplex networks could define a graph, depending on the clusters' dissimilarities (Rastin and Kanawati 2015); a search tree method was used to study the relevant layers for each community based on the subspace clustering method (Hanteer and Rossi 2019). A multiplex graph neural network algorithm was implemented to deal with multi-behavior recommendation problems based on online user behaviors (Zhang et al. 2020). A novel generative model was proposed to study the structure of the behaviors (layers) that could evolve in parallel on the same users (Basu et al. 2015).

Furthermore, multiplex networks are used to study the connections between users through multiple layers representing different types of relations in many complex fields like biological, social, and technological systems; a novel semi-supervised method was proposed to maximize the mutual information between users (Mitra et al. 2021). In this research, the model could pull the cluster-aware, node-contextualized graph and clusters across the layers of multiplex networks. The authors demonstrated the model's performance on many real-world multiplex networks and presented each case study's classification, clustering, visualization, and similarity search (Mitra et al. 2021). Finally, (Ding and Wang 2021) concluded that the experiments on real-world multiplex networks show that the significant role played by central users spread information across different layers and show higher accuracy for layer clustering compared with other methods. For this purpose, the novel in this research is to implement the multiplex networks method to enhance the discovery and interpretability of the focal structure set on social networks. Additionally, this contribution would help

highlight the influential sets of coordinating users' behavior, interactions, roles in disseminating conspiracy theories, harmful healthcare information, and fake news on social networks.

3 Research problem statement

The proposed research aims to implement the contextual focal structure analysis model into the social network analysis. Given the raw datasets from the online environment, the research problem statement is to implement a new model that utilizes the FSA 2.0 model and the multiplex network approach to reveal the users' activities, and the focal structure sets behavior in social networks. This approach involves different layers, including users' followers, mentions, retweets, URLs, and contexts in the form of participation layers in the solution procedure.

Moreover, this research presents essential ideas and should bring proper analysis to the significant questions like, How the CFSA model could help decision-makers understand the activities performed by the influential coordinating groups on social networks in real-time? Where implementing the traditional community detection methods is not sufficient to identify these coordinating groups, neither can reveal the contexts activities in one analysis? Is the proposed research would need an operational method and systematic multidisciplinary approaches? Is the CFSA model able to let the reader know what is happening between online users? What are the topics disseminated between online communities on social networks? What users and hashtags bridge different online communities? What are the reactions of the influential users to the famous/trending contexts spread on social networks? Furthermore finally, can the CFSA model help identify activities that will give birth, grow up, or disappear on/from online social networks?

This research should answer many other operational questions about online fake news spreaders, conspiracy theories, misinformation, and how to suspend this behavior on social networks.

4 Methodology

In this paper, we developed a contextual focal structure analysis approach (CFSA) that leverages hashtags as a context to improve the discovery and interpretability of focal structures in social networks. We can implement this idea using the multiplex network, where it could be generated by combining several participation layers, considering one layer to be the user's network based on mentions, replies, friends, and followers. The second layer could be the hashtag co-occurrence network. In addition, this research observes the interconnections based on users and hashtags relations, where it seeks to see users' connections in different communities. Still, as explained earlier, they share similar contexts and interests on social networks.

The CFSA model is designed to consider the nonlinear relationships of online users in the social network that generate different channels (layers) like shared topics, breaking news, and blogs, which can conceptualize the different layers of the multiplex network. In other words, the interconnection links between users and their activities generate the multiplex network, the union of the users–users layer, and the online activities like the hashtag–hashtag layer. Also, the users–users layer is a unimodular network that includes users' activities with other users, such as mentions, followers, and likes in the network. The second layer represents the context activities of the context in the social network, such as the hashtag–hashtag layer includes hashtags mentioned in other hashtags. Finally, the outcomes from the CFSA model are influential sets of users linked to other significant users and, at the same time, connected to the activities and contexts (like hashtags) shared by users (participation layer) on the social networks. For this purpose, we implemented FSA 2.0 model presented in Alassad et al. Jan. (2021), and the multiplex networks approach in Cozzo et al. (2018), as shown in Fig. 3.

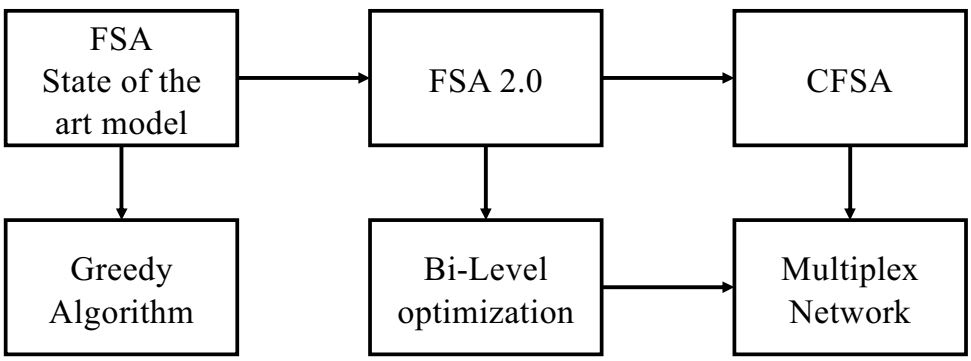


Fig. 3 Focal structure analysis modeling

5 Notations and representation

Consider a multiplex network $\mathcal{M} = (L, \boldsymbol{n}, B, M)$, where $L = \{1, 2, \ldots, m\}$ is an index set of layers that we call the user–user layer and the hashtag–hashtag layer. $\boldsymbol{n}$ is a set of nodes, and $B = (\boldsymbol{n}, L, R)$, where $R \subseteq \boldsymbol{n} \times L$ is a binary relation between the user–user and hashtag–hashtag layer. Let $(n, \alpha) \in R$ represents a statement of node n participating in layer α, where the ordered pair $(n, \alpha) \in R$ a node-layer pair and we say that the node-layer pair (n, α) is the representative of node n in layer α (user–user or hashtag–hashtag layers).

On the other hand, $M = \{G_\alpha\}_{\alpha \in L}$ is a set of graphs, let us consider graph G_F on R in which there is an edge between two node-layer pairs (n, α), and (m, α) only if $n = m$; that is, only if two edges in the graph G_β are incident on the same node $n \in \boldsymbol{n}$, which means that the two node-layer pairs represent the same node in different layers. We call G_F the coupling graph, in which it is formed by $n = |\boldsymbol{n}|$ disconnected components that are clique or isolated nodes. Each clique is created by all the representatives of a node in the layers, which we call the components of G_F supra-nodes.

Let us now also consider the graph G_l on the same nodes set R, and in which there is an edge between two node-layer pairs, $(n, \alpha), (m, \beta)$ only if $\alpha = \beta$; that is, only if the two edges in the graph $G_\beta \in M$ are incident on the same node n and layer $\alpha \in L$. We call G_l the layer graph. It is easy to realize that a graph is formed by $m = |L|$ separate components that are cliques.

Finally, we define the supra-graph $G_\mathcal{M}$ as the union of the layer-graph with the coupling graph $G_\mathcal{L} \cup M$. $G_\mathcal{M}$ has node-set R and edge set $\cup_\alpha E_\alpha \cup E_F$. $G_\mathcal{M}$ is the synthetic representation of Multiplex Network $\mathcal{M}$. It results that each layer-graph G_α is a sub-graph of $G_\mathcal{M}$ induced by $\boldsymbol{n}_\alpha$. Furthermore, when all nodes participate in all layer graphs, the multiplex network is said to be a fully aligned (Cozzo et al. 2018), and the coupling graph is made of n complete graphs of m nodes.

5.1 Modified adjacency matrix for CFSA

In general, the adjacency matrix of an unweighted and undirected graph G with N nodes in a $N \times N$ symmetric matrix $\boldsymbol{A} = \{a_{ij}\}$, with $a_{ij} = 1$, only if there is an edge between i and j in G, and $a_{ij} = 0$ otherwise. We can consider the adjacency matrix of each of the graphs introduced in the previous section. The adjacency matrix of layer graph G_α is $n_\alpha \times n_\alpha$ symmetric matrix $\boldsymbol{A}^\alpha = a^\alpha_{ij}$, with $a^\alpha_{ij} = 1$ only if there is an edge between (i, α), and (j, α) in G^α. We call them layer adjacency matrices.

Likewise, the adjacency matrix of G_β is an $n \times m$ matrix $\rho = p_{i\alpha}$, with $p_{i\alpha} = 1$ only if there is an edge between the node i and the layer α in the participation graph, i.e., only if node i participates in layer α. We call it the participation matrix. The adjacency matrix of the coupling graph G_F is an $N \times N$ matrix $\mathcal{L} = \{c_{ij}\}$, with $c_{ij} = 1$ only if there is an edge between node-layer pair i and j in G_F, i.e., if they are representatives of the same node in different layers. We can arrange rows and columns of $\mathcal{L}$ such that node-layer pairs of the same layer are contiguous, and layers are ordered. We assume that $\mathcal{L}$ is always arranged in that way. It results that $\mathcal{L}$ is a block matrix with zero diagonal blocks. Thus, $c_{ij} = 1$, with $i, j = 1, \ldots, N$ representing an edge between a node-layer pair in layer 1(user–user layer) and node layer pair in layer 2 (hashtag–hashtag layer) if $i < n_1$ and $n_1 < j < n_2$.

The supra-adjacency matrix is the adjacency matrix of the supra-graph $G_\mathcal{M}$. Just as $G_\mathcal{M}$, $\overline{A}$ is a synthetic representation of the whole multiplex $\mathcal{M}$. It can be obtained from the intra-layer adjacency matrices and the coupling matrix in the following way:

$$\overline{A} = A^\alpha \oplus_\alpha \mathcal{L} \tag{1}$$

where the same consideration as in $\mathcal{L}$ applies for the indices, we also define. $\boldsymbol{A} = \oplus \boldsymbol{A}^\alpha$, which we call the intra-layer adjacency matrix.

5.2 Workflow

In this section, we provide the technical intuition into our model; it consists of three main components (data collection, CFS sets discovery, and CFS sets validation and analysis) as presented in Fig. 4. Also, the details on each step of the CFSA model and the solution procedure are explained in this section.

Step 1 A list of contexts was generated to feed our Python API used to collect Tweets from the Twitter Environment. These contexts include different trending hashtags related to Twitter network events, as shown in Fig. 4.

Step 2 The Python API was set to overcome the limitation imposed by Twitter in collecting contexts and running in real-time over different periods.

Step 3 This step is to collect contemporary contexts from the Twitter network over time, where the Twitter API was designed to accept more recently posted tweets than the older ones. In addition, we used Python libraries like Scarpy ("Scrapy | A Fast and Powerful Scraping and Web Crawling Framework". Available: https://scrapy.org/) and Tweepy ("Tweepy". [Online]. Available: https://www.tweepy.org/) to collect a preset list of co-hashtags related to different events on the Twitter Network as shown in Appendix A. COSMOS research laboratory at UA Little Rock put lots of effort into a dedicated research group to retrieve a considerable amount of data related to COVID-19 misinformation spread and different social movements on Twitter connected to anti/pro-COVID-19 health regulations. The researchers collected data related to COVID-19 vaccines, the 2020 USA election,

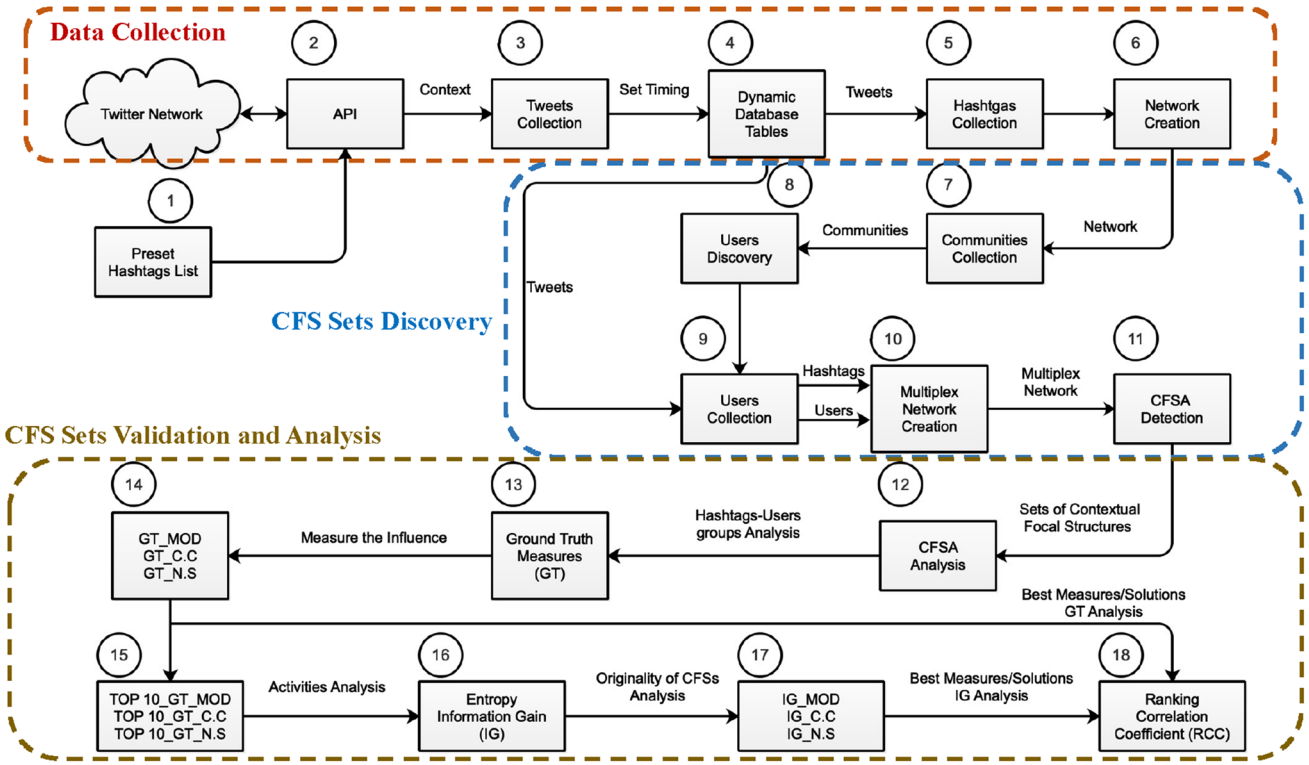


Fig. 4 Overall CFSA model structure. The model is divided into three main steps represented in the research methodology; the first is data collection, the second is CFS sets discovery and multiplex networks, and the third is CFS sets validation and analysis

ANTIFA, and other exciting datasets, including social activities like "Black Lives Matter" and "MAGA" on Twitter and YouTube ("COVID-19 MISINFO | Home Page" 2021). For this research, we used two different datasets shown in Table 1 (hashtag–hashtag, users–users, and users–hashtag networks) to evaluate the performance of the CFSA model.

Step 4 The data were retrieved in real-time, stored in different tables, and segmented into columns depending on the content to serve the requirements of this study, as presented in Fig. 4.

Step 5 We retrieved trending hashtags and related features such as retweets, mentions, and any tweets that include related content.

Step 6 Based on the collected information from (Step 5), we generated a unimodular hashtag–hashtag network; the co-occurrence hashtags (first layer) in the multiplex network. In this step, we consider the size of the network and the nature of the context used in our case studies.

Step 7 After generating the co-occurrence hashtags network, we measured the number of communities, using the modularity method (Newman 2004); this step would help to harvest those online users from the database that participated in contexts activities, posted, shared, and retweeted related contexts on Twitter as shown in (Step 8).

Step 8 Based on the features available and the communities, we combine a list of users and the hashtags collected in (Step 6) and (Step 7). This step would help generate users' metrics connected to the hashtag–hashtag network. This step was achieved by implementing proper queries that resulted in a bulk collection of users and other related information, like profile info, number of tweets, retweets, number of followers, geographic information, usernames, mentions, and

Table 1 Datasets retrieved from Twitter. Users–Hashtags network (UH), Users–Users (UU), Hashtags–Hashtags network (HH), Communities in the network(C), Modularity values (M), Average Clustering Coefficient values (ACC), Nodes (N), Edges (E

Network	UH		UU		HH		C UH	M UH	ACC nUH	Period
	E	N	E	N	E	N				
Trump Vaccine	19,843	9694	1876	9456	401	295	21	0.45	0.163	May2020-January 2021
Bill Gates	97,110	36,108	96,586	35,797	579	780	88	0.52	0.155	January2020-July 2020

the web link. This information should be enough to generate the interconnection network and serve the purpose of this research.

Step 9 Utilizing the data collected and stored in (Step 4) and the discovered users from the information retrieved in (Step 8); we identified sets of users involved in different activities, their followers, and links to generate the co-occurrence users network. This step helps develop the second layer in the multiplex network, the user–user network, where the outcome is a unimodular network including the users linked to other users without any context activities, as shown in Fig. 5.

Step 10 In this step, we generate the multiplex network; it is the union of the hashtag–hashtag layer in (Step 6) and the user–user layer as explained in (Step 9). This layer is about the interconnections between users and hashtags. To elaborate, we compare the enhancement achieved on the users' level in the CFSA model here versus FSA 2.0 model, as presented in Fig. 5. The centrality method used in the FSA 2.0 model is shown in Eq. 2; this equation is used to maximize the users' centrality values at the user level, where only a unimodular user–user network is used in the body of the model

$$max \sum_{i=1}^{n} (\delta_i) \tag{2}$$

where δ_i is the centrality value for user i, (degree, betweenness, closeness) values; n is the number of users in the unimodular network. However, in the CFSA model, Eq. 3 shows the modified centrality equation used at the user level. This equation measures the centrality values in the interconnection multiplex network ($L = 2$).

$$max \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\delta_i^{UU} \oplus \beta_{ij}^{UH} \hbar_j^{HH} \right) \tag{3}$$

where n is the number of nodes in the user–user layer UU. m is the number of nodes in hashtag–hashtag layer HH. δ_i^{UU} is the sphere of influence for users i in UU. $\oplus$ is the direct sum. $\hbar_j^{HH}$ is the number of j nodes in HH connected by an edge to user i in UU. Finally, β_{ij}^{UH} represents the interconnection between users and hashtags, where $\beta_{ij}^{UH} = 1$ if and only if node i in UU has a link with node j in HH, otherwise 0.

Step 11 We implemented the user–hashtag network into the CFSA model. This model can accept contexts, users, and

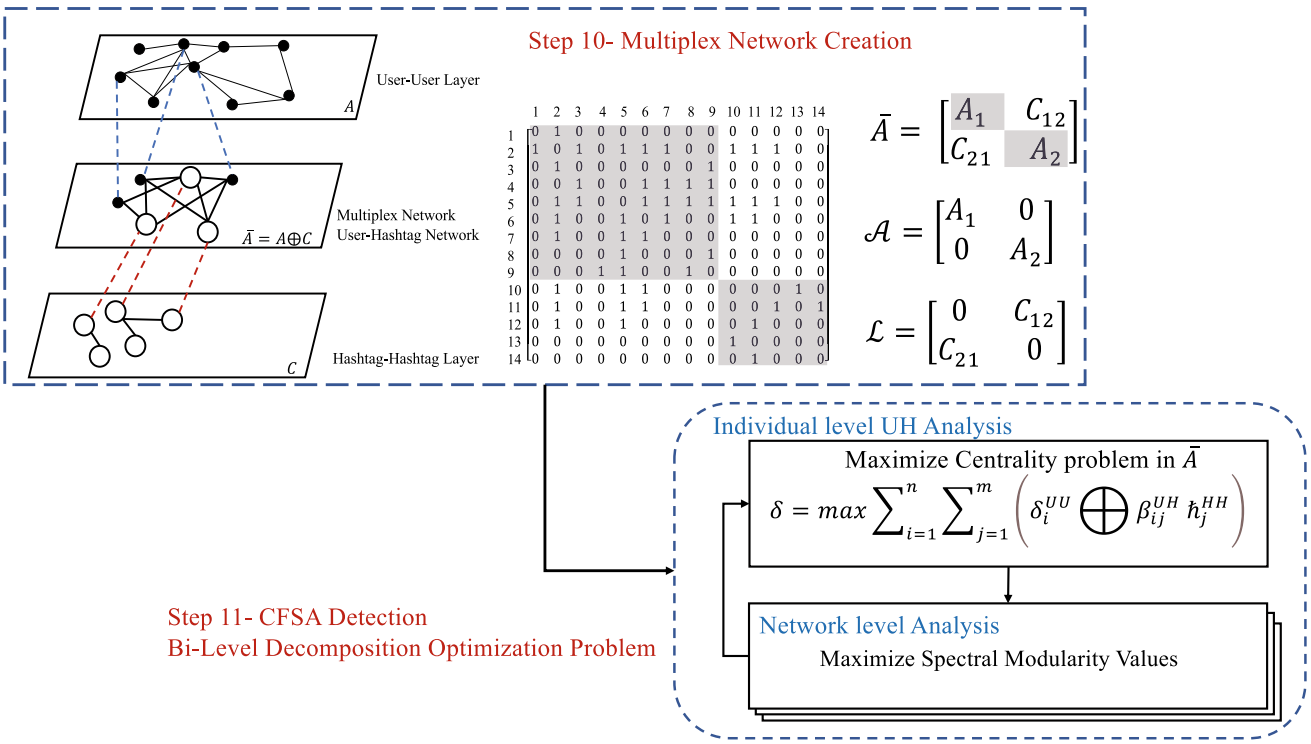


Fig. 5 The CFSA model takes an initial set of user–user Twitter network (1st layer) and hashtag–hashtag (co-occurrence) Twitter network (2nd layer) where $\boldsymbol{L} = 2$. Then, we generate the interconnection network called the user–hashtag layer or the coupling matrix $\bar{\boldsymbol{A}}$. The CFSA model utilizes a strategy of three phases: (1) the discovery of new activities between users–users in the first layer, (2) the discovery of activities between hashtag–hashtag activities (co-occurrence) in the second layer, and (3) the discovery of contextual activities using the users–hashtags in the third layer

the user–context links, where this combination represents the coupling matrix $\overline{A}$ shown in Eq. (1). The outcomes from this step are the smallest possible contextual focal structure sets, including influential users acting in different communities, linked to other influential users, and at the same time related to the contextual activities (hashtags). Our work in this domain successfully identified key sets of users that influence the maximum number of users in the network. The model combined two well-known social network analysis methods, namely centrality and modularity, to bridge the shortcomings of traditional community detection methods used in graph theory. The resultant combination is a bi-level linear optimization problem (Alassad et al. 2021b) to realize/observe user- and network-level interactions, as presented in Fig. 5.

Step 12 This step is to manually analyze the identified CFS sets, such as the size, number of users, number of edges, and hashtags in each CFS set.

Step 13 To validate the contextual focal sets and quantitatively measure their impacts in social networks, we will use the ablation method to calculate the focal sets' influence and power when each focal set is suspended from the network. This process exposes the campaign's dependence on each contextual focal structure's activities by determining the pockets that control the flow of information between users. This process also reveals each focal structure's activities and identifies information diffusion in social networks. In addition, we measure the quality of the CFS sets identified in (Step 11), where three ground truth (GT) measures are utilized to calculate the amount of influence any CFS set could generate in the entire structure of the network.

Depending on the amount of changes in the network and to evaluate the influence of the results in the network. The model would suspend each CFS set from the network ($G - CFS_i$), then recalculate the changes concerning the modularity values, the number of new communities, and the links between users in the network. Likewise, this step allows us to measure the importance (influence) of each CFS set in the network, where the model suspends all CFS sets and measures the changes in the network before and after suspending each CFS set from the network. In this step, the model would calculate three measures after each iteration and order the CFS sets in descending order based on the results. For this purpose, we implemented three measures as follows:

The first measure The model first utilizes the Newman–Girvan modularity method (Newman 2004; Clauset et al. 2004) to measure the general impacts each focal structure has on the network and monitor the changes in the communities after suspending each focal set in the network. Suspending the influential focal sets from the network will change the network's structure, disconnect many users, and cause other communities to disappear altogether from the network. In this step, we measured the amount of change through the modularity values after suspending CFS_i from the entire network. The ground truth modularity values are (GTMOD).

The second measure Ground truth clustering coefficient (GTCC) is the second method used to measure the impacts generated by suspending each focal set in the network to study the links changes between users after suspending the CFS_i set from the network. For this purpose, we utilized the clustering coefficient method (Zafarani et al. 2014) to observe the changes at the individual level.

The third measure Ground truth network stability (GTNS) is the third method used to validate the model's outcomes. GTNS is to measure the global impacts generated by suspending theCFS_i set in the network. This measure observes the change in the number of communities after suspending CFS_i from the network. This step talks more about the network and what it looks like after suspending any CFS_i from the network.

Step 14 In this step, we sorted the CFSs based on their influence in the network concerning the GT measures (Step 13). This step would narrow down the solutions and provide more flexibility in the analysis to present the GT values of all CFS sets.

Step 15 Select the top ten CFS sets from each GT measure (GTMOD, GTCC, and GTNS). To elaborate more, to select the top ten in GTMOD, we measured the changes in modularity values after suspending all CFS sets. We sorted the results in descending order where $G - CFS_{\underset{MOD}{x}} > G - CFS_{\underset{MOD}{y}} \cdots > G - CFS_{\underset{MOD}{z}}$. In this scenario, we select the top ten sets that maximized the network's modularity values, as shown in Sect. 5. The same procedure would apply to the other two GT measures (GTCC and GTNS).

Step 16 The entropy information gain (IG) is used to find out more about the structure of the top ten CFS sets. We used the IG theory to measure the originality of the top ten CFS sets, find the differences in the CFS sets' users, links, and context, and calculate how much information a CFS set provides in the solution. Shannon's model defines entropy for this step, as shown in Eq. 4.

$$\mathrm{IG}\left(\mathrm{CFS}_i\right) = -\sum_{i=1}^{\ell} P\left(\mathrm{CFS}_i\right) \log_2 P\left(\mathrm{CFS}_i\right) \tag{4}$$

Moreover, the idea with IG is that the more heterogeneous and impure a CFS set is, the higher the entropy values. Conversely, the more homogeneous and purer a CFS set is, the lower its entropy value. For this purpose, we used a Python code to measure IG values for the top ten CFS sets, where ℓ is the number of CFS sets identified in (Step 13), $P\left(\mathrm{CFS}_i\right)$ is the probability of selecting the target (root set) CFS_i set appearing in the entire solution. To elaborate more, let us

consider a dataset with N number of CFS sets, where $CFS_i \in N$ and nodes and edges where $(n_i, E_i) \in CFS_i$, where information gain value of target set $IG(CFS_i)$ is the probability of appearing $(n_i, E_i) \in CFS_i$ in set $CFS_{\substack{j \\ i \neq j}}$ as shown below.

To summarize, these steps represent the solution procedure of the CFSA model, where three levels are considered to complete and enhance the focal structure sets analysis in the social networks.

$$IG(CFS_i) = -\left(P(CFS_j)\log_2 P(CFS_j) + P(CFS_{j+1})\log_2 P(CFS_{j+1} + \cdots + P(CFS_n)\log_2 P(CFS_n))\right) \quad (5)$$

Step 17 To measure the IG values for the top ten sets based on modularity (IGMOD), IG based on clustering coefficient values (IGCC), and IG based on the network stability (IGNS) for the top ten CFS sets identified in (Step 15), a Python code used to implement Eq. 4. The results are IG values between [-1, 1] for each measure (IGMOD), (IGCC), and (IGNS).

Step 18 Ranking correlation coefficient (RCC) is to find the correlation between the outcomes in (Step 13) and (Step 17), respectively. This measure should discover the strength of a linear relationship links between the CFS sets based on GT and IG measures. Furthermore, RCC values support the analysis and find the most feasible solutions based on (Step 13) and (Step 17). For this step, we implemented three different experiments as follows:

Experiment 1 This step measures the RCC values between values in (Step 13); (GTMOD, GTCC, and GTNS).

Experiment 2 This step calculates the RCC values between values in (Step 17); (IGMOD, IGCC, and IG_NS).

Experiment 3 in this step, we measure the RCC values for IG values measured in (Step 17) versus the GT values measured in (Step 13).

5.3 FSA 2.0 model versus CFSA model

Throughout this research, we enhanced the discovery and interpretability of focal structures in social networks. In the state-of-the-art model, Şen et al. implemented a greedy algorithm to discover focal structure sets in the social network (Şen et al. Dec. 2016). Next, Alassad et al. developed a decomposition optimization model to find enhanced focal structure sets in the social networks (Alassad et al. 2019a).

In Fig. 6, we see two focal sets from a real-world Twitter network, the set on the left was identified via FSA 2.0 model, and then the set on the right was determined by the CFSA model. However, the set on the left shows only the users' activities; for example, node "balleralert" is connected to other nodes like "Freedom," where no more information is provided here. Likewise, the model cannot provide more information about other activities between users in the network and will limit the analysis.

On the other hand, the set identified through the CFSA model includes users' and communities' actions associated with more information about online contextual activities. For example, users like "balleralert" and "Nick" are active friends on Twitter; nevertheless, based on these results,

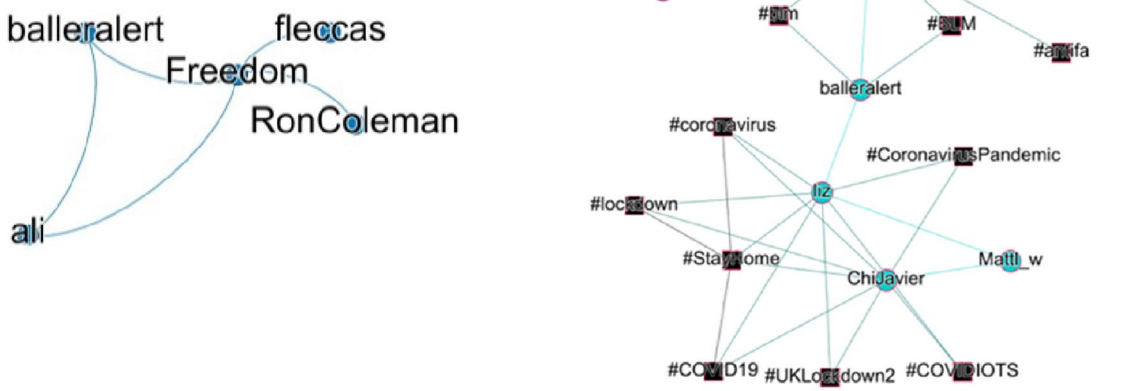


Fig. 6 FSA 2.0 versus CFSA model

these two accounts were supporting social movements like "#blm" and "#BLM." At the same time, we see a link between "balleralert" and "liz," where user "balleralert" occupies a central position in this set, and this link between these two users is a connection between two different communities (BLM supporters in the USA and the COVID-19 lockdown supporters in the UK). To elaborate more, we see a user like "balleralert" interested in the Black Lives Matter social movement, and user "liz" supported the COVID-19 safety-related hashtags and the second UK COVID-19 lockdown. Here we can illustrate the significance of CFSA modeling, which could represent more information about the connections and activities between different communities on Twitter.

Moreover, in Fig. 6, the set on the left is a clique of five users with no further information. Also, in the set on the left, no contextual information was included in the solution procedure, and the network was built only based on users–users with limited activities. For example, only user "balleralert" manages to survive in a CFS set identified by the CFSA model utilizing the multiplex network approach. In addition, the set's topology was enhanced due to the extra layer (contextual information) that was added to the solution procedure, where the CFSA model improves the examination and adds a new level of complication to the analysis.

Moreover, we compared the run time (R) complexity for both FSA 2.0 and CFSA models, where we run both models using a MacBook Pro with a 2.4 GHz 8-Core Intel Core i9 processor and 32 GB 2400 MHz DDR4 memory. Still, the R factor depends on the networks' density values, the number of nodes (n) in the network, and the number of layers used in the multiplex network. The run time of the FSA 2.0 model to execute this experiment using a unimodular network (users–users layer) was ($R < 9000$ s). However, the run time complexity of the CFSA model is $O(N \times \sum_l z^l_{max})$ where z^l_{max} denotes order or the number of layers in the multiplex network and N refers to the total number of nodes. The run times between FSA 2.0 and CFSA show that the latter algorithm has a slightly higher execution time. Comparison R between FSA 2.0 and CFSA model shows that the last experienced a slightly higher run time ($R < 16{,}000$ s). Nevertheless, the CFSA model accepts multiplex networks and the quality of the outcoming focal structure sets is better compared to FSA 2.0.

6 Results

This research is designed to show the benefits of the contextual focal structure analysis model that could increase the quality and enhance the discovery of the focal structure sets on social networks. The CFSA model should identify influential sets of users responsible for disseminating similar contexts on the Twitter network. These findings include influential users supporting popular hashtags and a better understanding of the information diffusion between communities on social networks, as presented in this section.

6.1 Trump vaccine network

The first case study implemented in this research was related to the Trump Vaccine dataset in Table 1. The CFSA model identified 187 CFS sets in the multiplex network (Users–Hashtags layer), where these sets are different in size, number of hashtags, user accounts, and network behavior. Additionally, these active/influential CFS sets include coordinating users and linked to other users who simultaneously shared similar contexts (hashtags). Likewise, these CFS sets disseminated various contents related to COVID-19 vaccines, the Trump Vaccine topic, and other related anti/pro-health-related content on Twitter.

Moreover, Table 2 shows three influential sets selected based on GT measures after suspending each CFS set from the network. For example, the CFS5 maximized the network segmentation (modularity values) based on the GTMOD measure, CFS27 minimized the connections (links) between users based on GTCC measure, and CFS187 minimized the network stability values based on GTNS measure to show the structure of the CFS sets in the model.

Furthermore, Table 2 shows the manual analysis and the activities of the CFS sets, observing "what is going on between online users?" in the most straightforward and smallest possible sets. For example, CFS5 includes 65 users and 21 hashtags disseminated on Twitter, where we observed four different communities with different behaviors on Twitter. In addition, this set contains influential users who share information and influence thousands of users on Twitter. For instance, the "#TrumpVaccine" hashtag was supported by a far-left sub-community, where this hashtag was linked to influential accounts such as "realDonaldTrump," "RepDLambor," and many other users on Twitter, as shown in Fig. 7 (A-CFS5). On the other hand, CFS5 set includes other sub-communities with users who showed utterly different opinions and interests (2nd sub-community); these users were disseminating content (hashtags) like "#BidenVaccine," "#GOPBetrayedAmerica," and "#PutinVaccine." Moreover, to describe the structure of the CFS5 set in-depth, Fig. 7 (A-CFS5) left side shows the spread of users (red dots) and the shared content (dark squares) at the structure of the network, where this set is considered as one of the top influential sets that include users from the different parts of the network and shared popular hashtags as mentioned earlier.

Likewise, the other CFS sets (CFS27 and CFS187) mentioned in Table 2 are presented in Fig. 7 (B-CFS27, C-CFS187), including influential users and Twitter

Table 2 CFS sets in complex social networks (Case study 1)

CFS Sets	Number of User	Number of Hashtag	Number of Edge	Communities (elements of different communities were identified as shown in the annotations)
CFS5	65	21	161	4 Far Right, Left, CDC, COVID, China Related
Contexts	#Trumpvaccine, #CovidVaccine, #TrumpVaccine, #WuhanVirus, #ChinaVirus, #COVID19, #vaccines, #vaccine, #Pfizer, #TrumpVirus, #trumpvaccine, #virus, #trumpvirus, #BidenVaccine, #PutinVaccine, #TrumpCrimeFamily, #trump, #OperationWarpSpeed, #Fauci, #BidenHarris2020, #GOPBetrayedAmerica			
CFS27	138	20	518	4 Far Right,Left, Right, Trump's Merchandise
Contexts	#USA, #WarpSpeed, #TrumpVaccine, #TrumpCasino, #TrumpSteak, #TrumpUniversity, #vaccine, #coronavirus, #SundayThoughts, #TrumpVirus, #trumpvaccine, #ETTD, #politics, #OperationWarpSpeed, #TrumpLied200KAmericansdied, #whitehouse, #covid19, #kayleighmcenany, #trump, #DonaldTrump			
CFS187	46	22	97	2 Far Left, far right
Contexts	#Biden, #coronavirus, #COVID, #COVID45, #KamalaHarrisVP, #LiarInChief, #SundayMorning, #SundayThoughts, #Trump, #TrumpCasino, #TrumpChaos, #trumpDementia, #TrumpLied200KDied, #TrumpSteak, #TrumpUniversity, #Trumpvaccine, #trumpvaccine, #TrumpVirus, #trumpwarpspeed, #TrumpVaccine#tRumpvaccine, #vaccine			

communities. To recap, the CFSA model enhanced the interpretability of the focal structure sets and went beyond the users–users connections in the analysis. Similarly, the initial results of the focal structure sets would observe and differentiate the context, users' interests, and the sub-communities involved in the information diffusion on social networks.

6.1.1 Ground truth measures

Three ground truth (GT) measures were employed to calculate the influence/importance of the CFS sets in the network. Sect. 4.3 (step 13) is utilized for actions like GTMOD, GTCC, and GTNS after suspending each CFS set from the network. Furthermore, when a CFS set is suspended from the network, it will change its structure, create new communities, and disconnect linked users from different parts of the network. For example, when the CFS5 set was suspended from the network (*G*-CFS5), it completely changed the network's structure and maximized the network's modularity values (GTMOD) from 0.45 to 0.711. Likewise, after the CFS187 set was suspended from the network (*G*-CFS187), this set minimized the stability (GTNS) of the network (maximized number of communities) values from 21 to 258 communities in the network. Similarly, suspending the CFS27 set from the network (*G*-CFS27) minimized the average clustering coefficient values (GTCC) from 0.173 to 0.132, as shown in Fig. 8.

In this matter, to evaluate the quality of the identified CFS sets, the model employed to suspend each CFS set and measure the changes in the modularity values, the changes in the number of communities, and the changes in the average clustering coefficient values, before and after suspending all CFS set from the network as shown in Fig. 9.

In summary, the CFSA model identified sets could influence the maximum number of users in the network, including the disseminated context in the form of hashtags, to overcome the limitations in the focal structure sets analysis. Meanwhile, the CFSA model identifies 187 CFS sets in this case study; then, we measured the importance of the sets in the network. We will focus on the top ten influential sets from each GT measure for the rest of the analysis. We will deliberate the information gained and the difference between the CFS sets, considering the size, number of users, hashtags, and links in each CFS set.

6.1.2 Information gain measures (IG)

To measure the information gain values as mentioned in Sect. 3.4 (Step 16), the IG method would help determine the amount of information each CFS set can deliver to the overall analysis. We selected the top ten influential sets based on the GT values. The top ten CFS sets based on GTMOD are (CFS5, CFS15, CFS10, CFS1, CFS184, CFS179, CFS9, CFS16, CFS19, CFS187), top ten CFS sets based on GT (C.C) are (CFS27, CFS26, CFS22, CFS28, CFS54, CFS104, CFS187, CFS36, CFS115, CFS24), and the top ten CFS sets based on GTNS are (CFS187, CFS27, CFS5, CFS15, CFS10, CFS1, CFS22, CFS3, CFS179, CFS9).

The process is to arrange the target CFS_i set in the model, then measure each CFS set's uniqueness (information gain) /(distance) concerning the target CFS_i set. Likewise, the model will measure the information gained against the abovementioned top ten sets. Figure 10 shows the IG values when the model arranged the target sets on (CFS5, CFS37, and CFS187), where CFS5 is highly dissimilar to CFS114, CFS27 is highly different to CFS82, and CFS187 is dissimilar to CFS44.

6.1.3 Ranking correlation coefficient values (RCC)

In this level of the analysis, the model would identify the correlated solutions and discovers the strength of the linear relationship between the CFS sets based on GT and IG values explained in sects. 5.1.1 and 5.1.2. In addition, the RCC

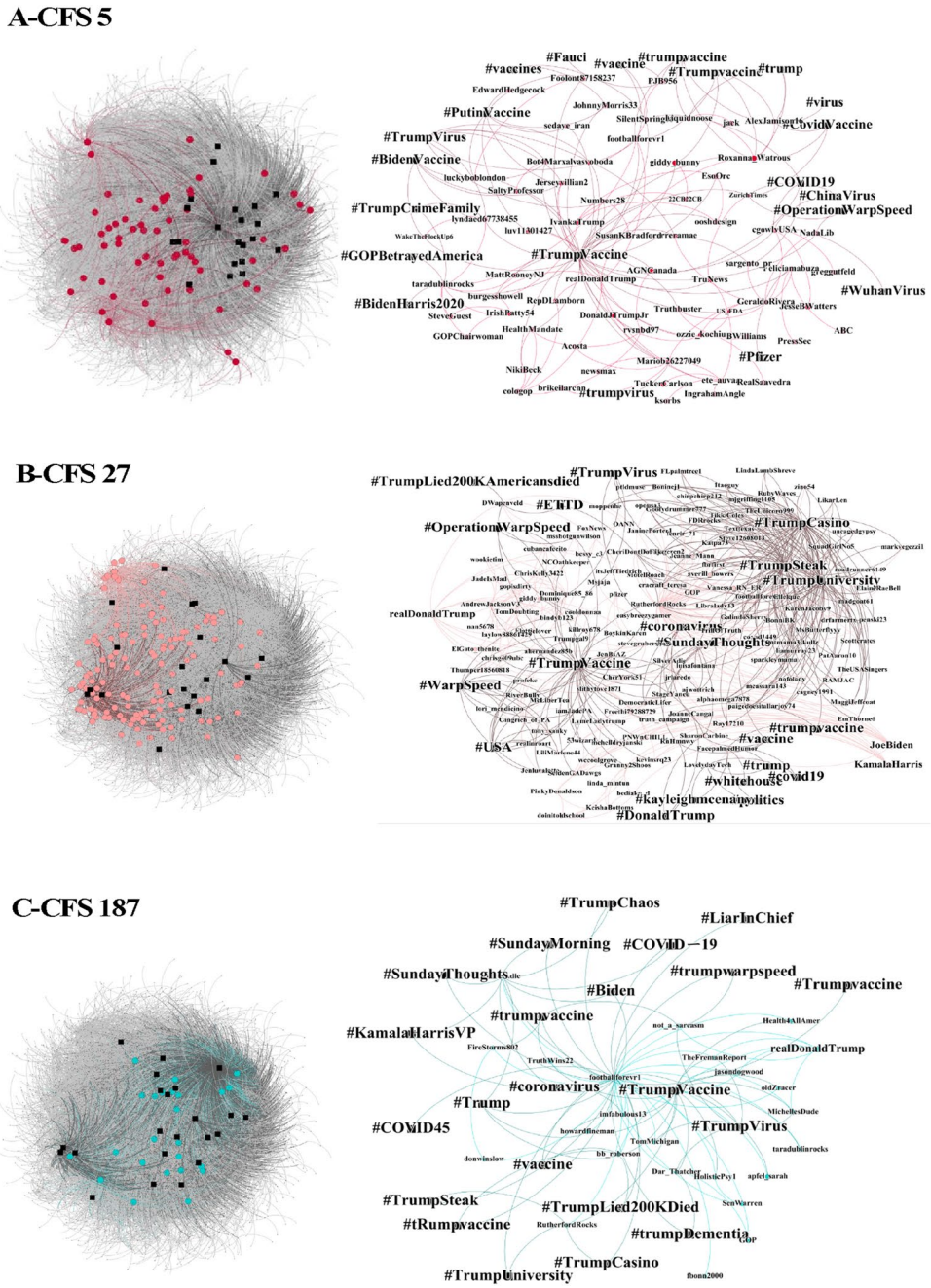


Fig. 7 CFS sets in social networks (Case study 1)

values are used to support the analysis of the focal structure sets, find the most feasible solutions, and help the decision-maker analyze the outcomes concerning the importance, most valuable information, and the CFS sets' structures in the network. We applied three experiments in Sect. 4.3 (Step 18) to find the correlated outcomes based on IG, GT, and IG vs. GT values.

Experiment 1 This step measures the RCC values between the 187 CFS sets based on GTMOD, GTCC, and GTNS values. This experiment shows that the CFS sets in GTMOD values are correlated with the results in GTCC values, where the RCC = 0.189, as shown in Fig. 11.

Experiment 2 This step measures the RCC values for the top ten CFS sets employed to find correlated solutions between CFS sets in the IG results. Thus, the correlated results were between the top ten CFS sets based on IGCC values and the top ten CFS sets based on IGNS measures, as presented in Fig. 11.

Experiment 3 This experiment measures the RCC values for the results in GT values vs. results in IG values. The outcome of this experiment includes ninety RCC values; we

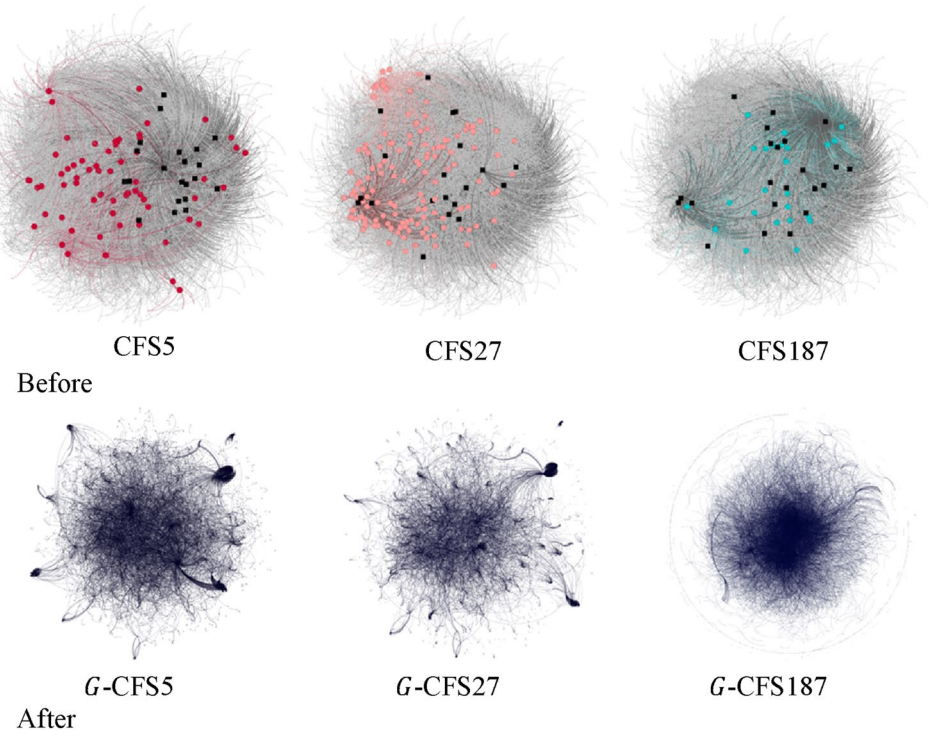
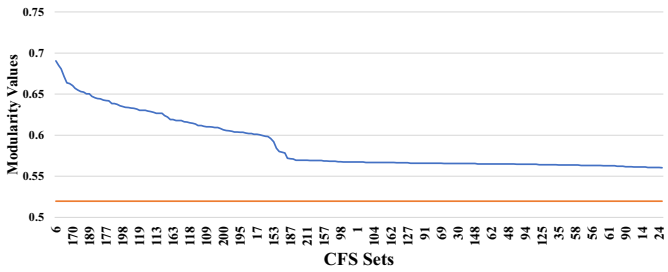
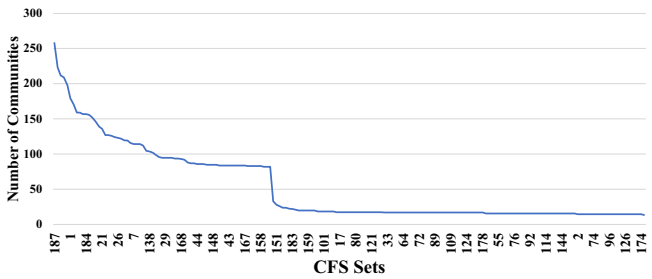


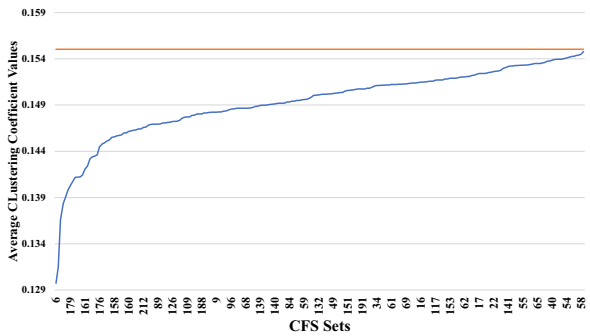
Fig. 8 Influence of CFS sets in social networks. These three CFS sets changed the structure of the network as we can observe the changes before and after suspending these three sets from the network



9.1: Modularity values changes.



9.2: Communities change in the network.



9.3: Clustering coefficient values change.

Fig. 9 Changes in the network after suspending CFS sets

measured the RCC values for the top ten based on IGMOD vs. the three GT measures (GTMOD, GTCC, and GTNS) values. Furthermore, to capture the overall correlation between these results, we calculated the average value for each RCC value, as shown in Fig. 11. The outcomes from this experiment offer the finest and the correlated solutions between IG and GT values in dark blue ink.

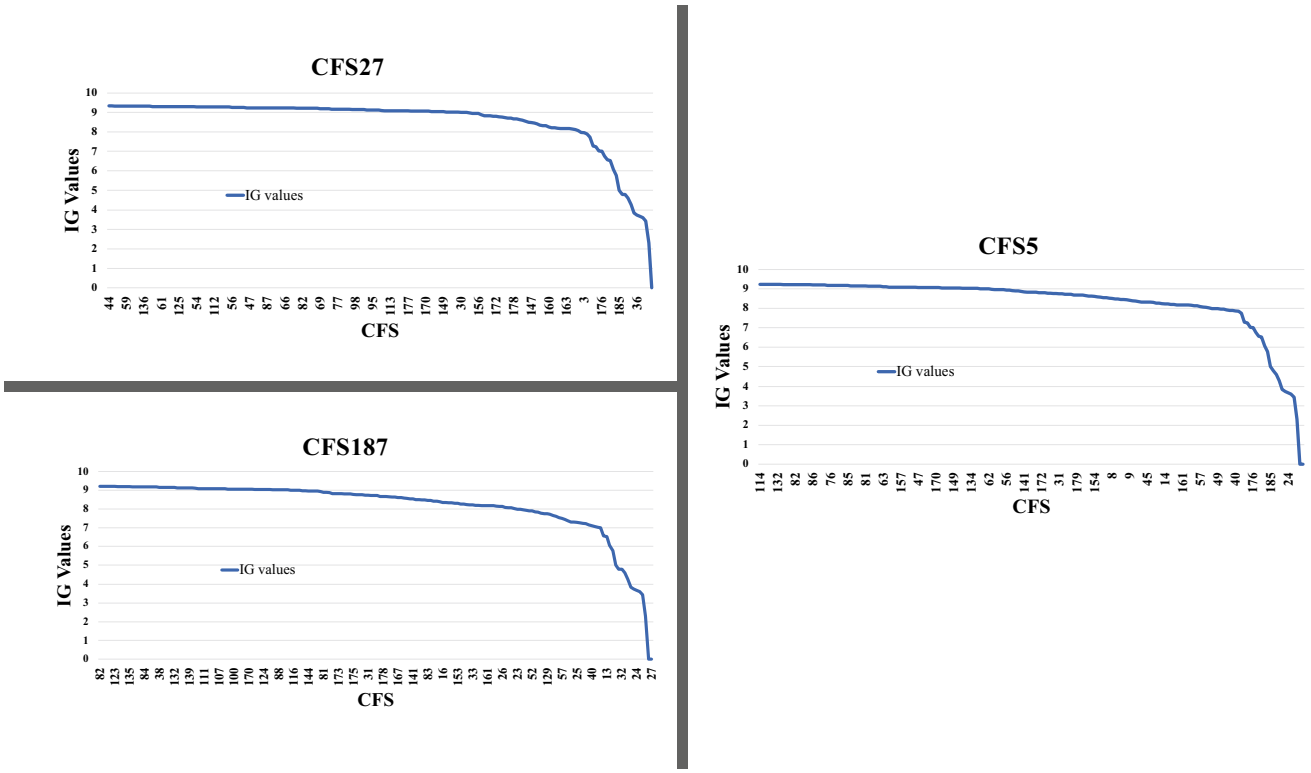
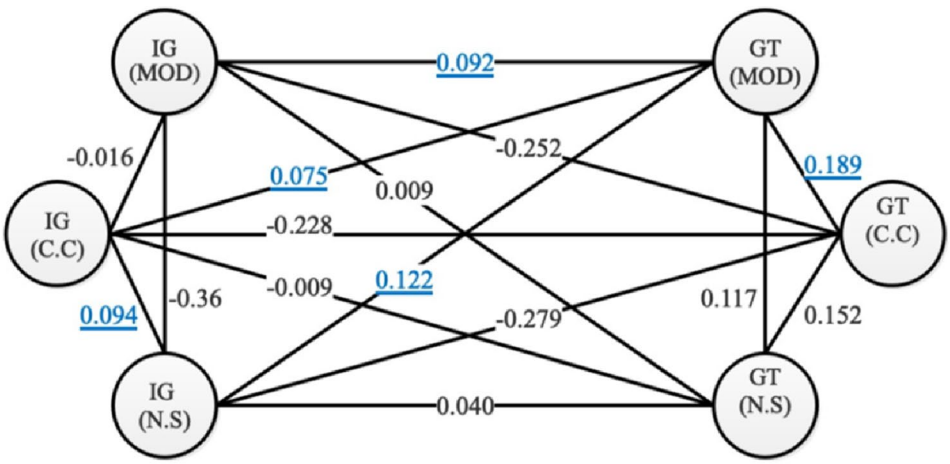


Fig. 10 CFS sets IG values

To recap the results from this case study, the CFSA model identified 187 CFS sets of contextual focal structure sets, including coordinating users in a complex Twitter dataset spreading information related to popular content (hashtags) such as "TrumpVaccine" and other COVID-19 vaccines' hashtags in 2021. We validated the results by measuring the influence of 187 CFS sets based on three GT measures. Then we studied the changes in the structure of the networks after suspending CFS sets from the network. Next, we measured the IG values for the top ten CFS sets to measure the distance between the CFS sets and the amount of information gained in the analysis. Finally, the model implemented the ranking correlation coefficient method (RCC) to find the semi-correlated solutions to be consistent with real-world scenarios ($0 < RCC < 0.3$), where the CFS set in the GTMOD experiment were correlated to CFS sets in IGMOD, IGCC, and IGNS.

Fig. 11 CFS set RCC values



6.2 Bill Gates network

The second case study implemented in this research was related to the Bill Gates dataset presented earlier. The CFSA model identified 218 CFS sets in the multiplex network (Users–Hashtags layer), where these sets are different in size, number of hashtags, user accounts, and network behavior. In addition, these active/influential CFS sets include coordinating users and linked to other users who shared similar contexts (hashtags) simultaneously. Likewise, these CFS sets disseminated content related to COVID-19 vaccines, Bill Gate's COVID-19 activities, and other related anti/pro-health-related content on Twitter.

Moreover, Table 3 shows three influential sets selected based on GT measures after suspending each CFS set from the network. For example, CFS6 maximized the network segmentation (modularity values) based on the GTMOD measure, CFS5 minimized the connections (links) between users based on GTCC measure, and CFS7 minimized the network stability values based on GTNS measure to show the structure of the CFS sets in the model.

Furthermore, Table 3 shows the manual analysis and the activities of the CFS sets, observing "what is going on between online users?" in the most straightforward and smallest possible sets. For example, CFS6 includes 42 users and 37 hashtags disseminated on Twitter, where we observed four different communities with different behavior on Twitter. In addition, this set contains influential users who share information and influence thousands of users on Twitter. For instance, hashtags like "#BillGates," "#BillGatesIsEvil," "#BillGatesVaccine," "#BillGatesBioTerrorist," "FakePandemic," and other related hashtags are disseminated on Twitter. Similarly, content related to Dr. Fauci appeared to be in this dataset, where few users were spreading hateful content like "#FauciFraud," "#FauciTheFraud," "#FireFauci," and #FireFauciNow," and many other users on Twitter as shown in Fig. 12 (A-CFS6). Moreover, to describe the structure of the CFS6 set in-depth, (A-CFS5) left side shows the spread of users (red dots) and the shared content (dark squares) at the structure of the network, where this set is considered as one of the top influential sets that include users from the different parts of the network and shared popular hashtags as mentioned earlier.

Likewise, the other CFS sets (CFS5 and CFS7) mentioned in Table 3 are presented in Fig. 12 (B-CFS5, C-CFS7), including influential users and Twitter communities.

To summarize, the CFSA model enhanced the interpretability of the focal structure sets and went beyond the users–users connections in the analysis. Similarly, the initial results of the focal structure sets would observe and differentiate the context, users' interests, and the sub-communities involved in the information diffusion on social networks.

6.2.1 Ground truth measures

Three ground truth (GT) measures were employed to calculate the influence/importance of the CFS sets in the network. Sect. 4.3 (step 13) is utilized for actions like GTMOD, GTCC, and GTNS after suspending each CFS set from the network. Furthermore, when a CFS is set suspended from the network, it will change its structure, create new communities, and disconnect linked users from different parts

Table 3 CFS sets in complex social networks (Case study 2)

CFS Sets	Number of User	Number of Hashtag	Number of Edge	Communities (elements of different communities were identified as shown in the annotations)
CFS 6	42	37	321	1 Anti-Bill Gates and Dr. Fauci groups
Contexts	#Agenda2030, #Agenda21, #ArrestBillGates, #Banned, #BillGates, #billgatesagainsthumanity, #BillGatesBioTerrorist, #BillGatesEvil, #BillGatesIsEvil, #BillGatesIsNotOurFriend, #BillGatesOfHell, #BillGatesVirus, #billgatesvirus, #ClintonCrimeFamily, #COVID19, #DepopulationAgenda, #ExposeBillGates, #FakePandemic, #FauciFraud, #FauciTheFraud, #FireFauci, #FireFauciNow, #freedom, #GatesForPrison2020, #NewsWars, #NoVaccine, #oms, #Plandemic, #Plandemic2020, #Scamdemic, #scamdemic, #scamdemic2020, #Vaccine, #vaccines, #VaccinesAreNotTheAnswer, #vacunas, #WWG1WGAWORLDWIDE			
CFS5	32	21	248	2 Anti-Bill Gates' activities, Covid19
Contexts	#Anonymous, #BillGates, #billgates, #billgates2020, #BillGatesBioTerrorist, #BillGatesIsEvil, #BillGatesVaccine, #BillGatesVirus, #Corona, #CoronaHoax, #coronavirus, #coronavirusuk, #COVID, #Covid_19, #COVID_19, #COVID19, #NWO, #Plandemia, #plandemia, #scamdemic, #WHO			
CFS7	6	23	82	2 Anti-Bill Gates and Dr. Fauci groups
Contexts	#BigPharma, #BillGates, #billgates, #BillGatesBioTerrorist, #BillGatesIsEvil, #BillGatesVirus, #CoronaVaccine, #coronavirus, #covid, #Covid_19, #COVID19, #COVIDIOTS, #Fauci, #Gates, #GAVI, #Hydroxychloroquine, #MSM, #NoMasks, #NWO, #vaccine, #vaccines, #Vaccines, #WHO			

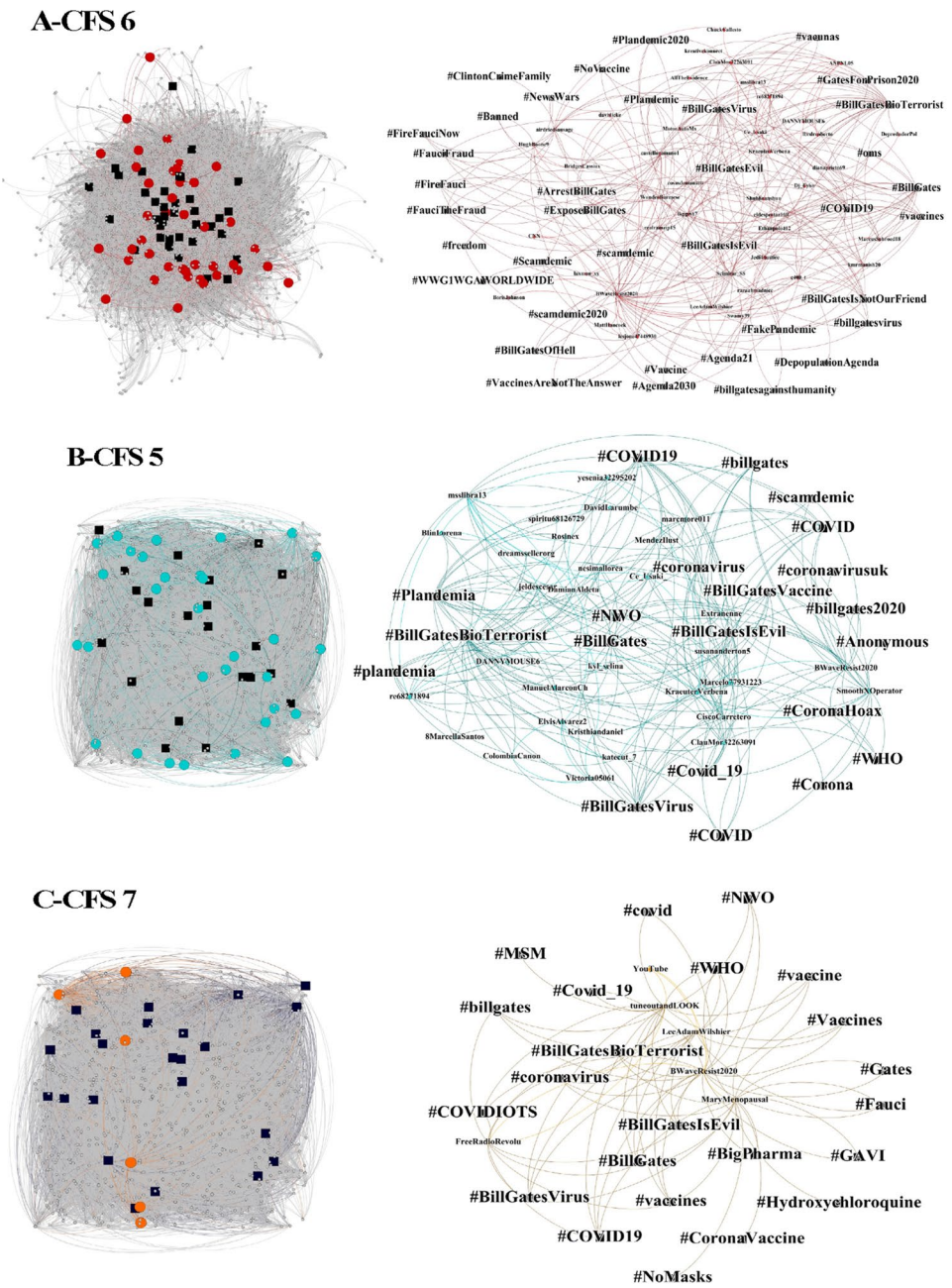


Fig. 12 CFS sets discovered from Bill Gates Network

of the network. For example, when the CFS6 set was suspended from the network (*G*-CFS6), it completely changed its structure and maximized the network's modularity values (GTMOD) from 0.52 to 0.7. In addition, when the CFS6 set was suspended from the network (*G*-CFS187), this set minimized the stability (GTNS) of the network (maximized number of communities) values from 88 communities to 2684 new communities in the network. Similarly, the same CFS set minimized the average clustering coefficient values (GTCC) from 0.155 to 0.129, as shown in Fig. 13.

Also, CSF#5 and CFS7 maximized GTMOD values from 0.52 to 0.68 after suspending these sets from the entire network. Correspondingly, the GTNS minimized when the model suspended CFS5 and CFS7, which increased the number of communities from 88 to 2101 and 1495, respectively. Finally, CFS5 and CFS7 minimized GTCC from 0.155 to 0.131 and 0.136, respectively.

Moreover, to evaluate the quality of the identified CFS sets, the model employed to suspend each CFS set and measure the changes in the modularity values, the changes in the number of communities, and the changes in the average clustering coefficient values, before and after suspending all CFS set from the network as shown in Fig. 14.

In summary, the CFSA model identified sets could influence the maximum number of users in the network, including the disseminated context in the form of hashtags, to overcome the limitations in the focal structure sets analysis. Meanwhile, the CFSA model identifies 218 CFS sets in this case study; then, we measured the importance of the sets in the network. We will focus on the top ten influential sets from each GT measure for the rest of the analysis. We will deliberate the information gained and the difference between the CFS sets, considering the size, number of users, hashtags, and links in each CFS set.

6.2.2 Information gain measures

This section is to measure the information gain values for the top ten influential sets versus the sets shown in Fig. 14. For this purpose, the top ten CFS sets based on GTMOD values are (CFS6, CFS7, CFS5, CFS8, CFS142, CFS202, CFS170, CFS2, CFS185, CFS173); the top ten CFS sets based on GTCC are (CFS6, CFS5, CFS7, CFS8, CFS142, CFS170, CFS179, CFS145, CFS186, CFS177), and the top ten CFS sets based on GTNS are (CFS6, CFS5, CFS170, CFS205, CFS7, CFS179, CFS200, CFS206, CFS8, CFS173). Furthermore, Fig. 15 shows the IG values when the model sets the target to be CFS5, CFS6, and CFS7 sets, respectively.

The process in this step of the analysis is to arrange the target CFS_i set in the model, then measure the uniqueness (information gain) /(distance) of each CFS set concerning the target CFS_i set. Likewise, the model will calculate the information gained against the above-mentioned top ten sets. Fig. 15 shows the IG values when the model arranged the target sets on (CFS6, CFS5, and CFS7), where CFS6 is highly dissimilar to CFS17, CFS5 is highly different to CFS27, and CFS7 is dissimilar to CFS61.

6.2.3 Ranking correlation coefficient values

In this step, the model would identify the correlated solutions and discovers the strength of the linear relationship between the CFS sets based on GT and IG values presented in sects. 5.1.1 and 5.1.2. Similarly, the RCC values are used to support the analysis of the focal structure sets, find the most feasible solutions, and help the decision-maker analyze the outcomes concerning the importance, most valuable information, and the CFS sets' structures in the network. We applied three experiments explained in Sect. 4.3 (Step 18) to find the correlated outcomes based on IG values, GT values, and IG vs. GT values.

Experiment 1 This step measures the RCC values between the 216 CFS sets based on GTMOD, GTCC, and GTNS values. This experiment shows that the CFS sets in IGMOD values correlate with the IGCC values results, where the RCC = 0.19, as shown in Fig. 16.

Experiment 2 This step measures the RCC values for the top ten CFS sets employed to find correlated solutions between CFS sets in the IG results. Thus, the correlated

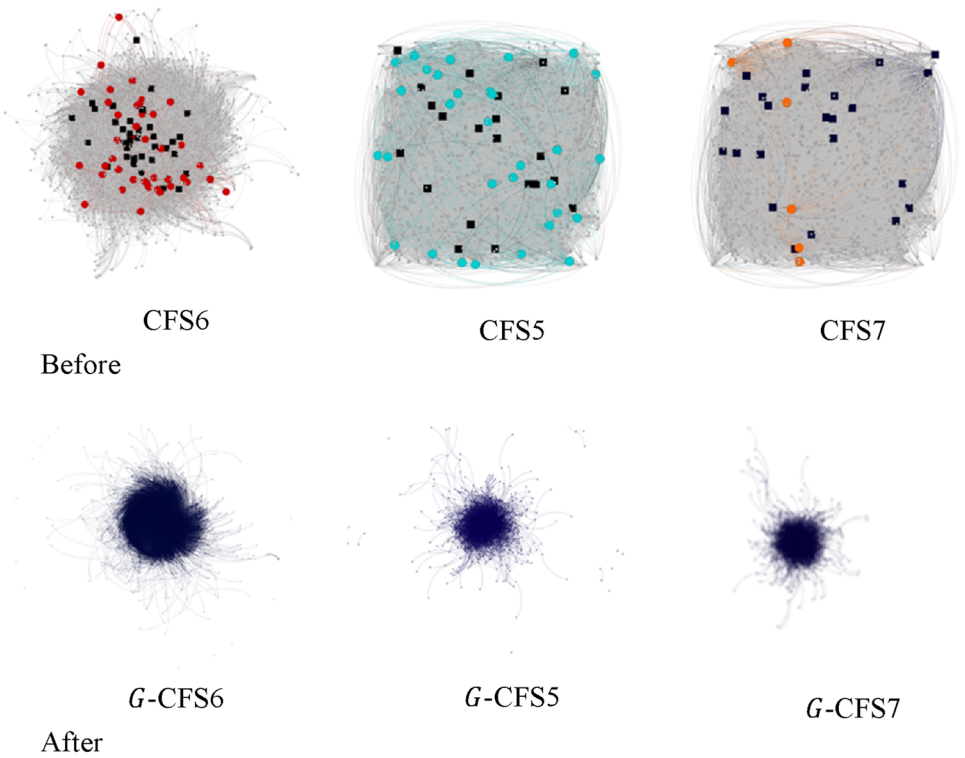
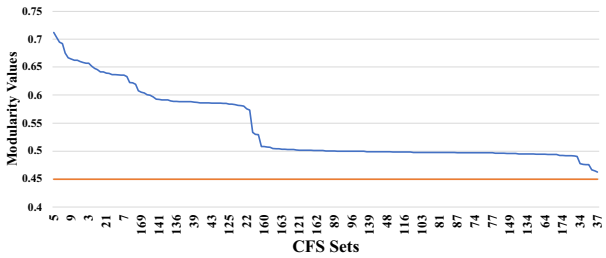
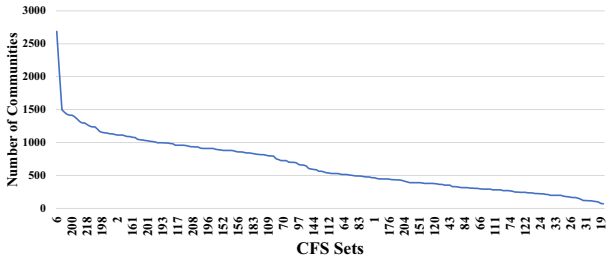


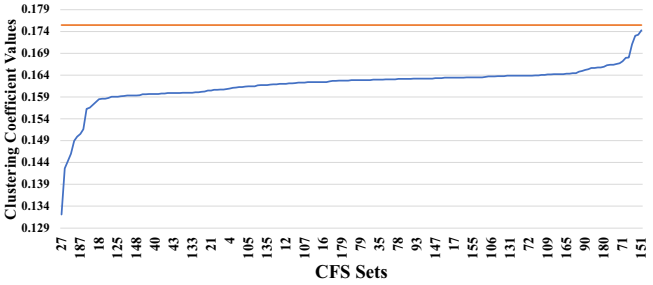
Fig. 13 CFS set influence in the network. These three CFS sets changed the structure of the network as we can observe the changes before and after suspending these three CFS sets from the network



14.1: Modularity change values.



14.2: Community change.



14.3: Clustering coefficient values change.

Fig. 14 Changes in the network after suspending CFS sets

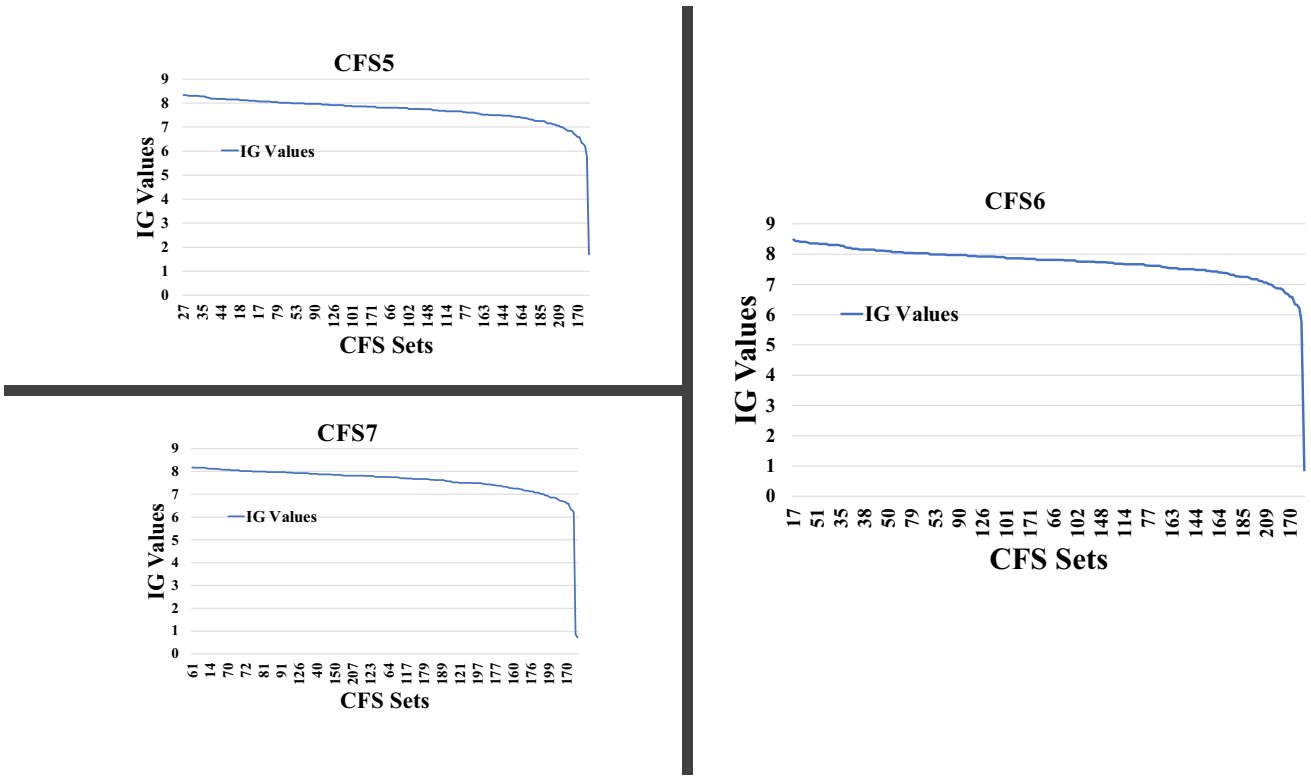


Fig. 15 CFS sets IG values

results were between the top ten CFS sets based on IGCC values and the top ten CFS sets based on IGMOD measures where RCC = 0.285, as shown in Fig. 16.

Experiment 3 This experiment measures the RCC values for the results in GT values vs. results in IG values. The outcome of this experiment includes ninety RCC values; we

measured the RCC values for the top ten based on IGMOD vs. the three GT measures (GTMOD, GTCC, and GTNS) values. Furthermore, to capture the overall correlation between these results, we calculated the average value for each RCC value, as presented in Fig. 16. The outcomes from this experiment show the finest and the correlated solutions between IG and GT values in dark blue ink.

In this case study, the CFSA model identified 216 CFS sets in the Twitter dataset related to Bill Gates' activities during the COVID-19 pandemic and other popular COVID-19 hashtags in 2021. We measured the influence of CFS sets based on three GT measures to validate the results. Then we studied the structure of the top ten CFS sets and measured the IG values. Finally, we used the RCC method to identify the correlated solutions from different measures; in this case study, GTMOD and GTCC are correlated, and IGMOD and IGCC are correlated. Finally, the top ten CFS sets are based on all three IG measures and GTNS.

6.3 Results' implications

Several theoretical and practical contributions are provided from this study that have been explained in this section.

6.3.1 Theoretical implications

The main theoretical implications we identified in this research are mentioned here. First, some efforts have explored the utilization of complex networks and the focal structure analysis characteristics in the detections of the contextual focal structures sets on social networks. Second, this study went beyond methods that focus on the users' followers, mentions, and retweets generated by users on social networks. In this study, we relaxed the analysis in a linear relationship, organized the added information, and helped interpret the users' contextual actions on social networks.

Third, this study highlights the bright and dark sides of the context activities of the coordinating groups on social networks. More information was revealed in this study; the reader would observe the users' interests, shared tweets or hashtags, or other information that will demonstrate their context behavior on social media.

6.3.2 Practical implications

This study provides several implications for practice. First, our study finds that multiplex networks and focal structure analysis models are positively related to revealing the coordinating groups' contextual activities and the information spread on social networks. It helps to analyze the users of social media's influence, the communities and coordinating groups' global influence, and then measure their impact on social networks. Further, this study suggests that these CFS sets should improve information literacy over time.

Second, this study verifies the performance of the proposed contextual focal structure analysis model in differentiating the contextual activities besides the focal structures on social media. Thus, social media platforms could apply the proposed characteristics to develop and implement screening tools for users' and communities' contexts activities on social networks.

7 Discussion and main findings

This research presents the contextual focal structure analysis (CFSA) modeling to reveal the online contextual activities of the coordination on social networks. We used the multiplex networks methods and the focal structure analysis model to expose the contextual activities of the influential sets of users in the form of users–hashtags on social networks. The multilayers or multiplex network methodology is utilized as a novel approach to revealing the interconnection layer

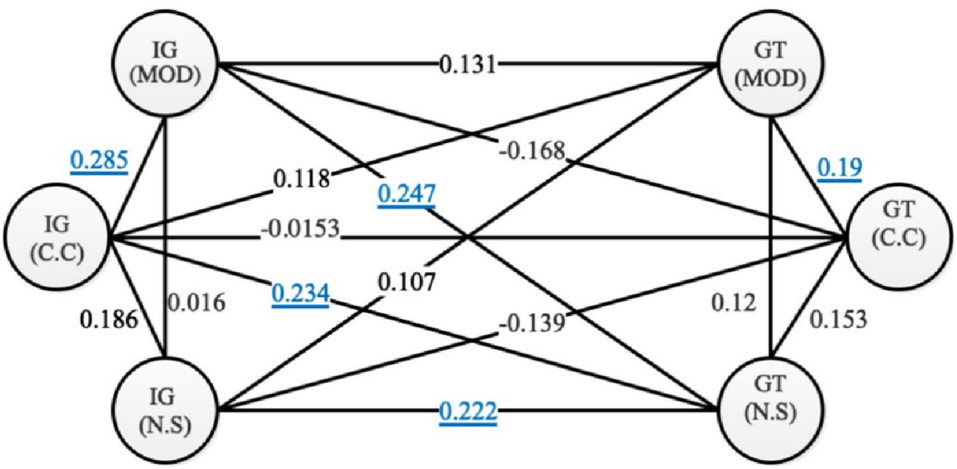


Fig. 16 CFS set RCC values

called the user–hashtag layer. Additionally, this layer consists of the union of edges between the user–user layer and hashtag–hashtag layer; the interconnection users–hashtag network includes the advantages of the communications between online users posting, sharing, and retweeting hashtags on Twitter. Moreover, we utilized different complex real-world social datasets collected from Twitter to measure the model's performance. These datasets related to events like COVID-19 vaccines, social movements like "Black Lives Matter" and "Make America Great Again" (MAGA), the USA 2020 election, and other related events we witnessed in 2020–2021. The CFSA model was able to identify the focal structure sets, the users' interests based on different layers, and the context activities in the social network. In addition, to evaluate the outcomes, we suspended every single CFS set from the entire network, then we utilized three different measures to calculate the changes in the network.

Furthermore, we measured the influence of each CFS set by re-calculating the changes in modularity values, clustering coefficient values, and network satiability across the network. Next, we used the other two measures to compare the CFS sets, where we calculated the information gain values for the top ten CFS sets for each ground truth measure. Finally, we used the ranking correlation coefficient factor to highlight the correlated results.

The research proposed here returns other conclusions. First, the contextual activities of the focal structures in complex social network analysis require multidisciplinary methods. This research enhanced the results of the focal structure sets to show the users–users links and their contextual interests in the same comment. In addition, including the context would avoid some uncertainties on the information disseminated by online spreaders on social networks.

The second conclusion is related to the influential sets of users; we found that the influential groups have links to popular trending contexts on social networks. Likewise, we witnessed the influential sets, including central users, were linked to different communities in social networks and able to control high information amount between communities on the network. This finding is consistent with the argument of previous studies (Şen et al. Dec. 2016) that the influential sets of users have higher centrality values and resources and can act in different parts of the network (Alassad et al. 2020). However, the influence of the central users is not significant for forcing all other users to accept all circulated contexts on social networks.

The third conclusion is related to the research's findings; we suggest that the outcomes from the CFSA model; witness the partisanship of the behaviors between the coordinating groups on social networks. This finding supports the hypothesis that an influential user's abilities to disseminate information on the social network are limited; it must be influential sets of users coordinating to spreading popular contexts or restricting the spread on social networks (Şen et al. Dec. 2016).

In the fourth conclusion of this research, the reader will find that the proposed CFSA model could reveal what is happening between users on social networks. These findings can help predict the dissemination of fake news, misinformation, flash mobs' radical behaviors, cyber-attacks, or the disappearance of critical and trending topics on social networks.

The last conclusion, the analysis in this paper, indicates that the outcomes from the contextual focal structure analysis model do have a high performance in identifying the influential focal structure sets and revealing the sets' contexts interests.

For future work, this paper provides a strong foundation; one area of future work is improving the CFS sets' legitimacy. For this purpose, we would like to implement different GT and IG measures to validate the results. This work would help in the validation level and could help implement other Entropy Information Gain theories.

Another area is in applying the model to dynamic social networks analysis. Since we cannot completely control the users joining communities or the coordinating groups, and participation in the information diffusion over time. Therefore, this work would help study the contextual activities of the dynamic coordinating groups over time.

The third area in the future work is implementing multi-cross online platforms' social networks analysis. Likewise, due to the complexities in the study of different social media platforms and the unpredictability in the users' behavior (where users have many accounts on other social media platforms), the CFSA model helps implement and examine cross-platform focal structures. The model presented here would accept more layers to include the cross-platform aspect of the problem domain, i.e., content shared on social media/communication platforms could be added to the CFSA multiplex network formalism. This extension would help study the contextual focal structure sets on multi-social platforms like (Facebook, Twitter, and Instagram) at the same time.

Appendix A

```
from tweepy import API
from tweepy import Cursor
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from textblob import TextBlob
import re
import sys
import tweepy
import json

ACCESS_TOKEN=" "
ACCESS_TOKEN_SECRET=" "
CONSUMER_KEY=" "
CONSUMER_SECRET=" "
class TwitterClient():
    def __init__(self, twitter_user=None):
        self.auth = TwitterAuthenticator().authenticate_twitter_app()
        self.twitter_client = API(self.auth)
        self.twitter_user = twitter_user
    def get_twitter_client_api(self):
        return self.twitter_client
    def get_user_timeline_tweets(self, num_tweets):
        tweets = []
        for tweet in Cursor(self.twitter_client.user_timeline, id=self.twitter_user).items(num_tweets):
            tweets.append(tweet)
        return tweets
    def get_friend_list(self, num_friends):
        friend_list = []
        for friend in Cursor(self.twitter_client.friends, id=self.twitter_user).items(num_friends):
            friend_list.append(friend)
        return friend_list
    def get_home_timeline_tweets(self, num_tweets):
        home_timeline_tweets = []
        for tweet in Cursor(self.twitter_client.home_timeline, id=self.twitter_user).items(num_tweets):
            home_timeline_tweets.append(tweet)
        return home_timeline_tweets
# # # # TWITTER AUTHENTICATER # # # #
class TwitterAuthenticator():
    def authenticate_twitter_app(self):
        auth = OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
        auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
        return auth
# # # # TWITTER STREAMER # # # #
class TwitterStreamer():
    """
    Class for streaming and processing live tweets.
    """
    def __init__(self):
        self.twitter_autenticator = TwitterAuthenticator()
    def stream_tweets(self, fetched_tweets_filename, hash_tag_list):
        # This handles Twitter authetification and the connection to Twitter Streaming API
        listener = TwitterListener(fetched_tweets_filename)
        auth = self.twitter_autenticator.authenticate_twitter_app()
        stream = Stream(auth, listener)
        # This line filter Twitter Streams to capture data by the keywords:
        stream.filter(track=hash_tag_list)
# # # # TWITTER STREAM LISTENER # # # #
class TwitterListener(StreamListener):
    """
    This is a basic listener that just prints received tweets to stdout.
    """
    def __init__(self, fetched_tweets_filename):
        self.fetched_tweets_filename = fetched_tweets_filename
    def on_data(self, data):
        try:
```

```
            print(data)
            with open(self.fetched_tweets_filename, 'a') as tf:
                tf.write(data)
            return True
        except BaseException as e:
            print("Error on_data %s" % str(e))
        return True
    def on_error(self, status):
        if status == 420:
            # Returning False on_data method in case rate limit occurs.
            return False
        print(status)
    def on_date (self, data):
        all_data          = json.loads(data)
        created_at        = all_data['created_at']
        source            =all_data['source']
class TweetAnalyzer():
    """
    Functionality for analyzing and categorizing content from tweets
    """
    def clean_tweet(self, tweet):
        return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ", tweet).split())

    def analyze_sentiment (self, tweet):
        analysis=TextBlob(self.clean_tweet(tweet))
        if analysis.sentiment.polarity>0:
            return 1
        elif analysis.sentiment ==0:
            return 0
        else:
            return -1
    def Tweets_to_data_frame (self, tweets):
        df=pd.DataFrame(data=[tweet.text for tweet in tweets], columns=['tweets'])
        df['id']= np.array([tweet.id for tweet in tweets])
        df['len']= np.array([len(tweet.text) for tweet in tweets])
        df['date']= np.array([tweet.created_at for tweet in tweets])
        df['source']= np.array([tweet.source for tweet in tweets])
        df['likes']= np.array([tweet.favorite_count for tweet in tweets])
        df['retweet']= np.array([tweet.retweet_count for tweet in tweets])
        #df['numerator']= np.array([tweet.numerator_at for tweet in tweets])
        #hash_tag_list = ['donald trum', 'hillary clinto', 'barak obama', 'bernie sanders']
        return df
if __name__ == '__main__'
    twitter_client = TwitterClient()
    tweet_analyzer = TweetAnalyzer()
    api = twitter_client.get_twitter_client_api()
    tweets = api.user_timeline(screen_name="realDonaldTrump", count=500)
    #print(dir(tweets[0].id))
    #print((tweets[0].retweet_count))
    df= tweet_analyzer.Tweets_to_data_frame(tweets)
    df['sentiment']= np.array([tweet_analyzer.analyze_sentiment(tweet) for tweet in df['tweets']])
    print (df.head(10))
```

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, N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research (FA9550-22-1-0332), U.S. Army Research Office (W911NF-20-1-0262, 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.

Declarations

Conflict of interest The authors have no relevant funding, financial or non-financial interessts to disclose.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors. The authors declare that they have no known ethics issue that could have appeared to influence the work reported in this paper.

References

Alassad M, Spann B, Agarwal N (2021) Combining advanced computational social science and graph theoretic techniques to reveal adversarial information operations. Inf Process Manag 58(1):102385

Alassad M, Spann B, Al-khateeb S, Agarwal N (2021a) Using computational social science techniques to identify coordinated cyber threats to smart city networks. Springer, Cham, pp 316–326

Alassad M, Hussain MN, Agarwal N (2021b) Comprehensive decomposition optimization method for locating key sets of commenters spreading conspiracy theory in complex social networks. Cent Eur J Oper Res 30(1):1–28

Alassad M, Agarwal N, Hussain M N (2019a) Examining intensive groups in youtube commenter networks. In: proceedings of 12th international conference, SBP-BRiMS 2019a, no. 12, pp. 224–233

Alassad M, Hussain M N, Agarwal N (2019b) 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

Alassad M, Hussain M N, Agarwal N (2020) Developing graph theoretic techniques to identify amplification and coordination activities of influential sets of users. In: Accepted in international conference on social computing, behavioral-cultural modeling, & prediction and behavior representation in modeling and simulation, 2020, pp. 192–201

Al-Khateeb S, Agarwal N (2014) Modeling flash mobs in cybernetic space: Evaluating threats of emerging socio-technical behaviors to human security. In: Proc. - 2014 IEEE Jt. Intell. Secur. Informatics Conf. JISIC 2014, vol. 7, no. 1, p. 328

Basu P, Sundaram R, Dippel M (2015) Multiplex networks. In Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, pp. 456–463

Bothorel C, Cruz JD, Magnani M, Micenková B (2015) Clustering attributed graphs: models, measures and methods. Netw Sci 3(3):408–444

Chen N, Liu Y, Chen H, Cheng J (2017) Detecting communities in social networks using label propagation with information entropy. Phys A Stat Mech Its Appl 471:788–798

Clauset A, Newman M E J, Moore C (2004) Finding community structure in very large networks Cond-Mat/0408187, vol. 70, p. 066111

Cozzo E, de Arruda GF, Rodrigues FA, Moreno Y (2018) Multiplex networks. Springer International Publishing, Cham

COVID-19 MISINFO | Home Page 2021. [Online]. Available: https://cosmos.ualr.edu/covid-19. [Accessed: 16-Jul-2021]

Ding C, Wang J (2021) Link reciprocity in directed multiplex networks. In: 2021 5th international conference on cloud and big data computing (ICCBDC), 2021, pp. 102–108

Demographics of social media users and adoption in the United States | Pew research center, 2021. [Online]. Available: https://www.pewresearch.org/internet/fact-sheet/social-media/. [Accessed: 15-Jan-2022]

Falih I, Kanawati R (2015) MUNA: a multiplex network analysis library. In: proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and Mining 2015, pp. 757–760

Guimerà R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. Phys Rev E - Stat Nonlinear, Soft Matter Phys 76(3):036102

Hagen L, Kahng AB (1992) New spectral methods for ratio cut partitioning and clustering. IEEE Tran Comput Des Integr Circuits Syst 11(9):1074–1085

Hanteer O, Rossi L (2019) The meaning of dissimilar: an evaluation of various similarity quantification approaches used to evaluate community detection solutions. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, 2019, pp. 513–518

B. Hu et al. (2020) Loan default analysis with multiplex graph learning. In: Proceedings of the 29th ACM international conference on information & knowledge management, pp. 2525–2532

Li Z, Liu J (2016) A multi-agent genetic algorithm for community detection in complex networks. Phys A Stat Mech Its Appl 449:336–347

Li Z, Yan F, Jiang Y (2015) Cross-layers cascade in multiplex networks. Auton Agent Multi Agent Syst 29(6):1186–1215

Li X, Zhou S, Liu J, Lian G, Chen G, Lin CW (2019) Communities detection in social network based on local edge centrality. Phys A Stat Mech Its Appl 531:121552

Luo D, Bian Y, Yan Y, Liu X, Huan J, Zhang X (2020) Local community detection in multiple networks. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 266–274

Magnani M, Hanteer O, Interdonato R, Rossi L, Tagarelli A (2021) Community detection in multiplex networks. ACM Comput Surv 54(3):1–35

Mitra A, Vijayan P, Sanasam R, Goswami D, Parthasarathy S, Ravindran B (2021) Semi-Supervised Deep Learning for Multiplex Networks. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, 2021, vol. 1, no. 1, pp. 1234–1244

Moradi M, Parsa S (2019) An evolutionary method for community detection using a novel local search strategy. Phys A Stat Mech Its Appl 523:457–475

Newman MEJ (2004) Detecting community structure in networks. Eur Phys J B - Condens Matter 38(2):321–330

Rastin P, Kanawati R (2015) A multiplex-network based approach for clustering ensemble selection. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, pp. 1332–1339

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

Scrapy | A Fast and Powerful Scraping and web crawling framework. [Online]. Available: https://scrapy.org/. [Accessed: 02-Jun-2022]

Tweepy. [Online]. Available: https://www.tweepy.org/. [Accessed: 25-Jun-2022]

Zafarani R, Abbasi MA, Liu H (2014) Social media mining: an introduction. Cambridge University Press

Zhang W, Mao J, Cao Y, Xu C (2020) Multiplex graph neural networks for multi-behavior recommendation. In: Proceedings of the 29th ACM international conference on information & knowledge management, 2020, pp. 2313–2316

Zou F, Chen D, Huang DS, Lu R, Wang X (2019) Inverse modelling-based multi-objective evolutionary algorithm with decomposition for community detection in complex networks. Phys A Stat Mech Its Appl 513:662–674