Evolution of Focal Information Spreaders in Dynamic Social Networks

Mustafa Alassad, Muhammad Nihal Hussain, and Nitin Agarwal

COSMOS Research Center, UA - Little Rock, Little Rock, AR, USA
{mmalassad, mnhussain, nxagarwal}@ualr.edu

Abstract. The infestation of coordinating communities of focal information spreaders on online social networks reached new limits over the past few years. Several methods have been applied to investigate and suspend these communities of information spreaders in static social networks. However, static networks do not capture the evolution and dynamics in these communities. In addition, researchers applied advanced operational methods such as game theory and evolving complex graphs to describe the realignment of the communities in dynamic social networks communities. Yet, these methods suffer from high complexity and need to include many variables into the modeling when the time dimension is added. For this purpose, in this research, we propose a systematic approach to examine the focal information spreaders and track their evolution in social networks over time. This novel approach incorporates the focal structure analysis model to identify the coordinating communities of information spreaders in social networks and the adaptation algorithm to study their change in dynamics over time. We evaluated our findings against a real-world dynamic Twitter network. This network was collected from Saudi Arabian women's Right to Drive campaign coordination in 2013. Our approach enables observing, predicting, tracking, and measuring the coordination among communities of spreaders over time. Likewise, this approach investigates and illustrates when the information spreaders escalate their activities, where do they concentrate their influence, and what coordinating communities of spreaders are more proactive than other sets in the network over time.

Keywords: Dynamic Social Networks, Focal Structure Analysis, Adaptation Algorithm, Betweenness Centrality, Modularity Method, focal information spreaders.

1 Introduction

Millions of people around the world use online social network (OSN) platforms such as Facebook, Twitter, and Instagram to communicate, shop, trade, advertise, book flights and catch up with relatives, friends, and co-workers momentarily. The widespread use of such platforms allowed many users, communities, and agencies to share public announcements and breaking news to influence maximum users quickly. However, in the past few years, most OSNs hosted, and suffered from unprecedented campaigns spreading misinformation, disinformation, conspiracy theories and fake news everywhere. These damaging campaigns deviated the decent use of OSNs into a dark path and damaged their reputations seriously. For example, misinformation and fake news spreaders damaged numerous economic systems around the world [1], influenced political and election campaigns [2], and recently the volatility of the stock markets [3]. Furthermore, the stock markets' volatility and fluctuations in the price of GameStop (GME) stock is a perfect example demonstrating that the coordinated campaigns on social networks have an impact on the real world. The campaign, now being termed as OccupyWall 2.0, started on Reddit and quickly gained traction, leading hordes of redditors to buy and sell the stock in coordinated fashion [3]. Likewise, a group on Twitter organized an armed movement against COVID-19 lockdown in Michigan state in May 2020 [4]. In response to these ongoing problems in social networks, several researchers have attempted to identify and suspend the coordinating of focal information spreaders on OSNs. These studies use traditional community detection methods such as centrality at the user level and the modularity method at the groups levels to identify these coordinating focal sets [5]. Nevertheless, these studies aggregate temporal local and global social interactions into one static graph ignoring the social networks' dynamic aspects [6].

However, the focal information spreaders are active communities in social networks and their behavior changes over time. For example, the fake news spreaders' influence could grow exponentially over time or disappear/shrink from the network after they have served their purpose. Likewise, these information spreaders are able to change in the terms of size and space from one time period to another, which makes the analysis a complex and intrinsically dynamic process. To investigate such NP-hard problems [7], the main contributions and challenges within this research are as follows:

- Identify and track the focal information spreaders in evolving social networks.
- Study and illustrate the influence of the focal information spreaders in dynamic social networks.

For the purpose of this research and to overcome the above challenges, the main objective in this research is to introduce a systematic approach that integrates the focal structure analysis model [8] to identify the focal sets spreading information in dynamic social networks, and the adaptation algorithm [9] to observe the behaviors of the focal information spreaders over time. In addition, this research proposes a way to predict, investigate, and suspend the focal information spreaders to the maximum number of users in different parts of the OSNs.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 describes the proposed methodology. Section 4 explains the experimental results. Section 5 is the research conclusion.

2 Literature Review

In this research, modeling the focal sets' behavior and their growth in dynamic OSNs is the point of interest. Until now, most of the community detection methods used to identify focal sets into static OSNs.

Şen et al. [10] introduced focal structure analysis in OSNs, the authors proposed a greedy algorithm to identify hidden focal structures responsible for spreading fake news in static social networks. Alassad et al. [11] presented a bi-level centrality-modularity model to examine intensive groups of co-commenters spreading fake news in a static YouTube channel. In this research the authors explored hidden intensive groups and ranked them for further investigations. Authors in [2] studied the key information spreaders in a complex social network by using a bi-level decomposition optimization method. In an extended study, Alassad et al. [1] used computational social science techniques to identify coordinated cyber threats to smart cities infrastructure networks. In this research, the authors identified intensive sets of aggressors, measured their power by utilizing the deviant cyber flash mob detection method. Authors in [8] implemented a comprehensive decomposition optimization model for locating key sets of commenters spreading conspiracy theory in static OSNs. Authors in [12] studied computational social science techniques to identify coordinated cyber threats to smart city static infrastructure networks. However, the methods mentioned in this section applied only to static social networks, where in this research we focus on the behavior of focal information spreaders in dynamic networks.

Moreover, many scholars investigated the community detection in dynamic social networks. The study proposed in [13] applied a game theory method to measure the agent's utilities over time, this method applied to capture regular dynamic communities. Dakiche et al. [14] identified two types of growth for a regular community in dynamic networks, first they defined the community diffusion growth, is when a community attracts new members through ties to existing members. The second definition is non-diffusion growth is when individuals with no prior ties become numbers themselves. Authors in [15] predicted lifespan of a regular dynamic community based on a consistent set of structural features extracted from the most important features to predict its lifespan. Authors in [16] applied similarity function to match the regular dynamic communities from one time step to another to detect changes such as from, merge, split, dissolve, and survive. Bródka et al. [17] were able to model a classifier to discover events between time steps, where each sequence was able to describe community evolution, and each sequence consists of several preceding community sizes and events which serve as input for classifiers. However, these methods applied to study and cluster the regular dynamic communities, but our goal in this research is to study the behavior of focal information spreaders in dynamic social network and present their influence over time.

3 Research Methodology

The main objective in this research is to design a systematic model that integrates the focal structure analysis model with the complexity of the time dimension to identify and track sets of focal information spreaders in evolving real-world networks.

Consider a temporal online social network G on a set of S snapshots $S = \{1,2,\dots,n\}$ is $G = \{G_1, G_2, \dots, G_n\}$ and time $T = \{t_1, t_2, \dots, t_n\}$, where $G_i = (V_i, E_i, T_i)$ represents the snapshot S_i with nodes $|V_i|$, and edges $|E_i|$ at time t_i. Each snapshot G_i is used to represents a connected social network at time t_i. Using Dynamic (Temporal) decomposition optimization method, given G_i, the objective function and the goal of this research is to find set K_{G_i} of focal sets where $K_{G_i} = \left\{k_{1G_i}, k_{2G_i}, \dots, k_{jG_i}\right\}$ that can influence the maximum number of users and maximize the sparsity in G.

3.1 Focal Structures Analysis in Dynamic Social Networks

There are various studies applied to identifying the focal sets responsible for misinformation, and fake news spread in static social networks as mentioned in section 2. The aim for the focal structure analysis is to identify hidden influential sets of users that can influence maximum number of users, mobilize crowds, and participate in different communities in different parts of the OSNs. Authors in [2] and [11] applied models to maximize the users' centrality values and the networks' modularity values to identify such hidden groups in complex static OSNs. However, in this research, the focal structure analysis (FSA) model is extended to identify the focal information spreaders in each snapshot of a dynamic OSNs as presented in Figure 1. In other words, the FSA model will identify sets K_{G_i} in snapshot $G_i \in G$, for the selected time t_i.

Moreover, to study dynamic OSN and track evolution of focal information spreaders as the network changes, we utilized the adaptive algorithm proposed by Nguyen et al. [9] as shown in Figure 1. In this algorithm, all $K_{G_i} \in G_i$ identified by FSA model and all sets $k_{jG_i} \in K_{G_i}$ are mapped to all other snapshots in $G_i \in G$ utilizing the adaptation algorithm to measure the changes, growth, and activities values in K_{G_i} over time. The resultant focal information spreaders consist of influential members having strong communication/links with each other members of the network, and their communication haven't broken as the network structure evolves over time.

Using the adaptive algorithm helps to avoid the use the recalculation methods to repeatedly measure all instances and changes in every snapshot of the complex dynamic OSNs. Also, using the recalculation methods to track the evolution of the network over time could be simple. But these methods include several disadvantages, such as the expensive execution time, trap in the local optimal solutions, and getting the same reactions to tiny changes to inactive local parts in the dynamic OSNs.

On the other hand, the adaptive algorithm provides the ability to track qualitative and quantitative changes generated by k_{jG_i} to influence the entire network over time. The main advantages for using the adaptive algorithm are due to being less computationally expensive, less time consuming method to obscure the difficulty of continuously recomputing variables such as focal structure sets from scratch, and this method can observe the local OSNs events and illustrate the significant transformation of the focal information spreaders over a long duration in dynamic OSNs.

3.2 Validation and Verification

To validate the focal information spreaders and quantitatively measure their impacts in dynamic social networks, we will use various methods to calculate the focal sets' influence and power when each focal set is suspended from the network at any designated time frame. It also reveals information to the researcher about where, when, and what focal sets are more active than others in the dynamic social networks. The results validation and verification methods used in this paper are as follows:

- First, we utilized the Newman-Girvan modularity method [18] to measure the general impacts of each focal structure has on the network. In addition, this method used to monitor the changes in the communities after suspending each focal set in the dynamic network as shown in Figure 1.

- The depth-first search and linear graph algorithm [19] is employed is to measure the local impacts generated by suspending each focal information spreaders in the dynamic network. The model can measure the number of disconnected users from each snapshot after suspending each focal set from the network.

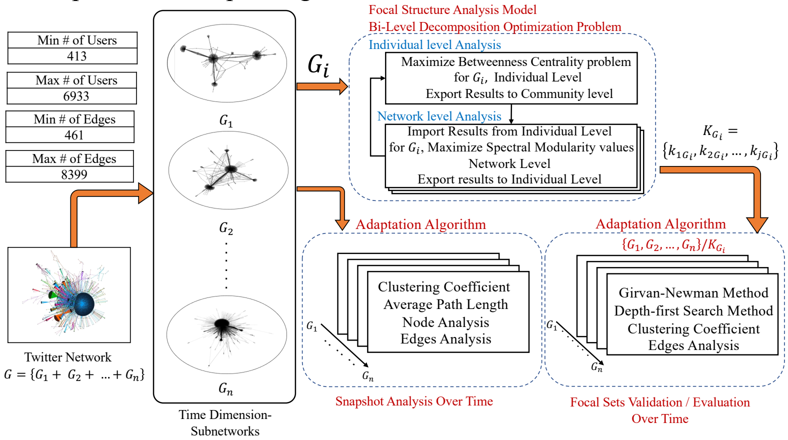

Figure 1: Overall model structure for conspiracy theories analysis in dynamic social networks.

- The adaptation algorithm [9] is used to observe the impacts generated by suspending the identified focal information spreaders over time. Implementing this algorithm will help to observe the changes in the communities and reduce the computation time by adapting the identified focal information spreaders into other snapshots.
- And finally, a dynamic Twitter network presented in Figure 2 was implemented to validate the model's performance.

4 Experimental Results

For the purpose of this research, the model will predict and track the growth in the focal sets in each snapshot following the three steps of analysis explained in section 3. The observation applied to a Twitter dataset related to the Saudi Arabian women's collective action of the ''Oct26Driving'' Campaign network. The structure of the network consists of max (min) number of users and edges as presented in Figure 1, and structured into 20 days (snapshots) as shown in Figure 2.

4.1 Dynamic Twitter Network

Figure 3 illustrates the Twitter network growth with respect to the number of users and links created over time. Likewise, the reader can observe the campaign was getting more popularity after snapshot # 3, on Oct. 13th as shown in the Figure 3 left side, where the number of users increased and then started to bend the curve on and after Oct. 28th, few days after the campaign day Oct. 25th.

In Figure 3 right side, we can observe the users' communications behavior over time, where all users went into massive activities and frequent actions to expand their influence on the Twitter network after Oct. 13th. The clustering coefficient values were maximized on Oct. 20th few days before the campaign day Oct. 25th.

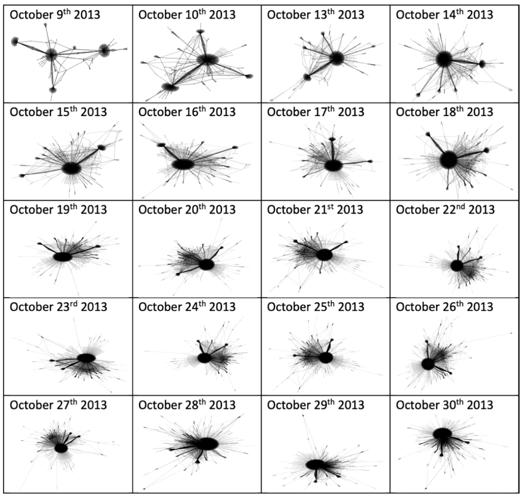

Figure 2: Dynamic Twitter Network for women's Activities in Saudi Arabia 2013.

In addition, the average path length values were minimized to reduce the distance between users and to maximize the spread of information before Oct. 25th. Moreover, the average path length values minimized on Oct. 20th, which suggests the increase in the users' activities few days prior to Oct. 25th. Also, we see the users' activities values were reduced on and after Oct. 28th.

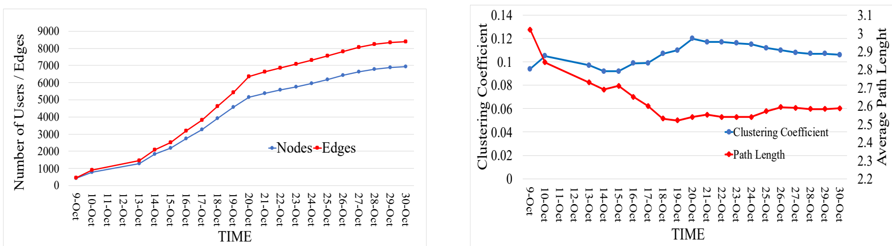

Figure 3: Users and links changes (left side) and Users' linking behaviors (right side) in an evolving social network. These values are considered as the original input snapshots to evaluate the activities in the dynamic network.

4.2 Focal Structure Analysis in Dynamic networks

The first step was to implement the FSA model on Oct. 9th includes only (413 users, 461 edges). The presumption for this study, Oct. 9th is when the campaign started spreading the agenda on the Twitter network publicly (the beginning of the data set), and the users began to influence other users on the network extensively. The advantages are for the purpose of the FSA model analysis, were we to allow the model to tracing and analyzing the focal information spreaders' communications and growth in the Twitter network over time. Also, selecting an early snapshot will help to validating the model's predictability feature, where the stakeholders can observe the focal information spreaders in the very beginning of the campaign. In addition, this process is a systematic method for blocking the focal information spreaders rather than suspending random influential users from the network. Furthermore, the early detection would help to track the focal sets evolutions and observe when they would merge with other communities, disappear or disconnect from the network.

The FSA model initially identified $K_{G_1} = 13$ focal information spreaders hidden in G_1, where these focal sets consist of influential users acting in different communities on the Twitter network. For example, Figure 4 presents the network on Oct. 9th before and after suspending focal set # 5 from the network. As presented, suspending this small focal set, (includes only 35 users), from the connected and highly dense network shown in Figure 4 (left) to go into complete sparse and disconnected scatters network as presented in Figure 4 (right). Likewise, we can confirm that the users lost their communications, disconnected from other users, and stopped information spread to other users and the entire network on Oct. 9th.

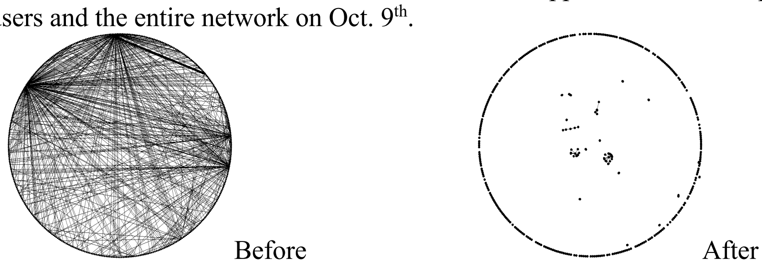

Figure 4 : Original Oct. 9th OSN (left). Snapshot of the OSN after suspending focal # 5 (right).

4.3 Validation and Evaluation in Dynamic Social Networks

The adaptation algorithm is utilized to validate the model's outcomes and track the focal sets' growth over time. The model will adopt and map the focal sets identified in K_{G_1} to measure the changes in other snapshots in G.

4.3.1 Clustering Coefficient and Modularity Values Over Time

Clustering coefficient method provides robust information on users' linking behaviors with other individuals in social networks [5]. Hence, the focal sets producing huge number of activities in different parts of the networks, suspending each focal set should disconnect large number of links (edges) and increase the users' sparsity in each snapshot. In other words, suspending any focal set should minimize the clustering coefficient values dramatically in compare to the values presented in Figure 3. Figure 5 left side shows the changes in the users' linking behaviors after suspending each focal set in K_{G_1} from all other snapshots in G. As presented in Figure 5 left side focal sets # (4,5, and 6) were able to decrease the clustering coefficient values, scattered the big complex communities into small powerless groups and ruined the users' connectivity with other users in the entire network over time.

Moreover, the modularity method implemented to measure the regular communities in each snapshot before and after suspending the focal structures [5]. Suspending any focal set should impact the entire structure of the network [1]. In other words, eliminating any focal set should sparse the network into smaller communities and maximize the modularity values at all snapshots in G. Figure 5 right side shows the changes in the modularity values after suspending each focal set in K_{G_1} from all twenty snapshots over time compared to compared to values in Figure 3. In addition, results show that suspending focal sets # (4, 5, and 6) increased the sparsity in the dynamic network.

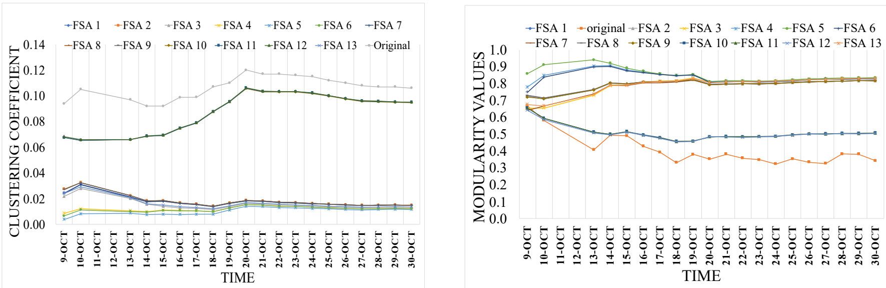

Figure 5: The activities of the users and communities in dynamic social network.

4.3.2 Network's Edges and Users' Connectivity Over Time

In this section, we show the changes in the number of users and edges after suspending each focal information spreaders. The adaptation algorithm illustrates a significant decrease in the users' connectivity after suspending focal information spreaders form each snapshot in G. In other words, focal sets in K_{G_1} occupying central positions in the network's structure to spread information all across the network. Moreover, Figure 6 left side validates the level of growth in the focal sets over time, where large number of edges disappeared after suspending each focal set. For example, suspending focal sets # (4,5, and 6) decrease the remaining number of edges in each snapshot in G compared to number of edges shown in Figure 3.

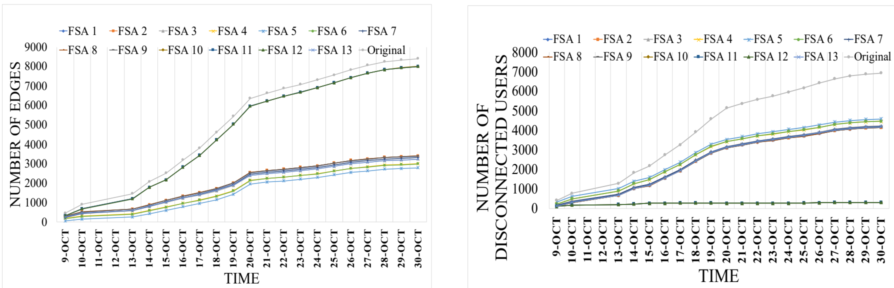

Figure 6: Changes in the connectivity of the Users and edges on dynamic social network.

In addition, Figure 6 right side shows the development on users' connectivity's after suspending each focal set from the evolving Twitter network, where the communities can grow, disappear, merge, and split from the network at any time frame. Likewise, the results recorded a massive increase in the number of disconnected users and communities after suspending each focal set in K_{G_1}. Moreover, suspending focal sets K_{G_1} identified on Oct. 9th helped to reduce the information spread to maximum number of users in all other snapshots as presented in Figure 6 right side. For example, suspending focal sets (4,5, and 6) maximized the number of disconnected users compared to original number of users connected to the network.

5 Conclusion and Discussion

In this research, we studied the capabilities of the focal sets in dynamic OSNs, where the focal structure analysis model used to identify the focal information spreaders were able to influence and transfer massive amount of information to the entire network over time. In addition, the adaptation algorithm utilized to observe the growth of the focal information spreaders over time, where (suspending) any focal sets should (minimize) maximize the clustering coefficient values, (minimized) maximize the number of edges between users, (maximize) minimize the sparsity and modularity values, and (maximize) minimize the number of disconnected users in dynamic OSNs. In addition, this research proposed a systematic and simplified method to investigate the evolution of the focal information spreaders over time, predict, and block the information spread in an evolving Twitter network. Finally, this research was able to overcome the complexities in measuring and tracking the focal information spreaders in dynamic social networks.

For future work, we would study alternative methods for the nodes' removal and to improve the effect of the user suspension from the network, where authors in [20] developed a modularity vitality method to calculate the exact change in modularity. Such research needs more investigation with respect to the focal structure analysis and the robustness of the network as mentioned in [21] and [22].

Acknowledgment

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), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

[1] M. Alassad, B. Spann, and N. Agarwal, “Combining advanced computational social science and graph theoretic techniques to reveal adversarial information operations,” Inf. Process. Manag., vol. 58, no. 1, p. 102385, Jan. 2021.

[2] M. Alassad, M. N. Hussain, and N. Agarwal, “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, 2019, pp. 41–54.

[3] “Robinhood, Reddit CEOs to Testify Before Congress on GameStop.” [Online]. Available: https://www.investopedia.com/robinhood-reddit-ceos-to-testify-in-congress-on-gamestop-gme-5112714. [Accessed: 16-Feb-2021].

[4] “Coronavirus: Armed protesters enter Michigan statehouse - BBC News.” [Online]. Available: https://www.bbc.com/news/world-us-canada-52496514. [Accessed: 29-Aug-2020].

[5] R. Zafarani, M. A. Abbasi, and H. Liu, Social Media Mining: An Introduction. Cambridge University Press, 2014.

[6] S. Wijenayake, “Understanding the dynamics of online social conformity,” in Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, 2020, pp. 189–194.

[7] C. J. H. Mann, Systems Thinking – Creative Holism for Managers, vol. 33, no. 8. 2004.

[8] M. Alassad, M. N. Hussain, and N. Agarwal, “Comprehensive decomposition optimization method for locating key sets of commenters spreading conspiracy theory in complex social networks,” Cent. Eur. J. Oper. Res., pp. 1–28, Feb. 2021.

[9] N. P. Nguyen, T. N. Dinh, Y. Shen, and M. T. Thai, “Dynamic social community detection and its applications,” PLoS ONE, vol. 9, no. 4. Public Library of Science, p. 91431, 10-Apr-2014.

[10] F. Şen, R. Wigand, N. Agarwal, S. Tokdemir, and R. Kasprzyk, “Focal structures analysis: identifying influential sets of individuals in a social network,” Soc. Netw. Anal. Min., vol. 6, no. 1, p. 17, 2016.

[11] M. Alassad, N. Agarwal, and M. N. Hussain, “Examining Intensive Groups in YouTube Commenter Networks,” Proc. 12th Int. Conf. SBP-BRiMS 2019, no. 12, pp. 224–233, 2019.

[12] M. Alassad, B. Spann, S. Al-khateeb, and N. Agarwal, “Using Computational Social Science Techniques to Identify Coordinated Cyber Threats to Smart City Networks,” Springer, Cham, 2021, pp. 316–326.

[13] H. Alvari, A. Hajibagheri, and G. Sukthankar, “Community detection in dynamic social networks: A game-theoretic approach,” in ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2014, pp. 101–107.

[14] N. Dakiche, Y. Slimani, F. B. S. Tayeb, and K. Benatchba, “Community evolution prediction in dynamic social networks using community features’ change rates,” in Proceedings of the ACM Symposium on Applied Computing, 2019, vol. Part F147772, pp. 2078–2085.

[15] N. Dakiche, F. Benbouzid-Si Tayeb, Y. Slimani, and K. Benatchba, “Tracking community evolution in social networks: A survey,” Inf. Process. Manag., vol. 56, no. 3, pp. 1084–1102, May 2019.

[16] M. Takaffoli, R. Rabbany, and O. R. Zaïane, “Community evolution prediction in dynamic social networks,” in ASONAM 2014 - Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2014, pp. 9–16.

[17] P. Bródka, P. Kazienko, and B. Kołoszczyk, “Predicting group evolution in the social network,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, vol. 7710, pp. 54–67.

[18] L. Donetti and M. A. Muñoz, “Detecting network communities: A new systematic and efficient algorithm,” J. Stat. Mech. Theory Exp., no. 10, pp. 1–8, 2004.

[19] R. Tarjan, “Depth-first search and linear graph algorithms,” SIAM J. Comput., vol. 1, no. 2, pp. 146–160, 1972.

[20] Magelinski, Thomas, Mihovil Bartulovic, and Kathleen M. Carley. "Measuring Node Contribution to Community Structure With Modularity Vitality." IEEE Transactions on Network Science and Engineering 8.1 (2021): 707-723.

[21] Requião da Cunha, Bruno, Juan Carlos González-Avella, and Sebastián Gonçalves. "Fast fragmentation of networks using module-based attacks." PloS one 10.11 (2015): e0142824.

[22] Holme, Petter, et al. "Attack vulnerability of complex networks." Physical review E 65.5 (2002): 056109.