Resilience and Node Impact Assessment in YouTube Commenter Networks Leveraging Focal Structure Analysis

Md Monoarul Islam Bhuiyan*, Shadi Shajari*, Nitin Agarwal*†
*COSMOS Research Center, University of Arkansas at Little Rock, Arkansas, USA
†International Computer Science Institute, University of California, Berkeley, California, USA
e-mail: {mbhuiyan, sshajari, nxagarwal}@ualr.edu

***Abstract*—Communication networks play a pivotal role in shaping information dissemination across social media platforms. Identifying influential groups or key players within these networks is essential for understanding how information flows and spreads. YouTube, as the leading video-sharing platform, offers a vast and dynamic environment for such studies. Our extended research centers on Focal Structure Analysis (FSA), aiming to identify core commenter groups within 35 YouTube channels discussing the Indo-Pacific region. By analyzing a dataset containing 308,890 videos, 726,078 commenters, and 1,536,284 comments, we apply two distinct FSA methods, namely FSA 1.0 and FSA 2.0, to detect influential network structures. We further evaluate the impact of these structures using network resilience metrics, including flow robustness and the giant component ratio. Our findings indicate that removing key focal structures results in a more fragmented and sparse network, significantly impairing information flow. This suggests that these core commenter groups act as critical bridges, facilitating communication and enhancing the cohesion of the network. By extending our prior work, this study offers deeper insights into the mechanisms of information spread on YouTube, providing a more comprehensive understanding of the platform's commenter dynamics.**

Keywords-Focal Structure Analysis; Social Network Analysis; YouTube; Network Resiliency.

I. Introduction

With the rise of social media platforms and their sophisticated recommendation algorithms, several aspects including content creation and sharing, news consumption, community engagement, societal influence, narrative propagation [1] and many other activities have gained wide popularity. This rapid adoption has become possible due to massive user engagement over content, driven by semiotics [2]. Every day, enormous amounts of information are generated through these platforms. While this rapid growth plays a pivotal role in the data sources for researchers, it is also crucial to find the best actionable knowledge from these data sources. Additionally, extracting actionable insights has widely been researched through the topology of complex social networks. As of 2024, YouTube is the second-most popular social media platform, the number one video-sharing platform globally, and available in over 100 countries and 80 languages its prominence in its user base has become streamlined due to its users massive engagement (views, comments, likes, shares, subscriptions, etc.) over the actual content [3]. Among these, YouTube's comment section provides a platform for constructive discourse, enabling viewers to share insights and directly connect with content creators. Despite this, the public discussion space can often lead to negativity and unproductive comments, which in turn can impair the user experience.

This paper conducts a comparative analysis of two distinct versions of Focal Structure Analysis namely FSA 1.0 [4] and FSA 2.0 [5], which is a social network analysis methodology designed to identify core sets of commenter groups within the co-commenter network of YouTube channels [6]. Initially, this study compares the outcomes of these two FSA approaches to evaluate their effectiveness in extracting relevant focal structures. Following this comparison, it addresses two research questions:

- **RQ1:** How do focal structures within a complex social network impact its resilience, as measured by network resilience metrics?
- **RQ2:** How much does each node in a particular focal structure contribute to the overall robustness?

By exploring the significance of these core groups and their impact on network resilience, this study aims to provide insights into the structural dynamics and robustness of social networks.

The rest of the paper is organized as follows. Section II reviews existing studies on identifying focal structures, detecting authoritative and community approaches, and measuring network resiliency metrics. Section III outlines the methods used for collecting data in this study. Section IV describes the experimental methodologies applied, while Section V presents the findings of our research. Finally, Section VI summarizes the study and suggests directions for future research.

II. Related Work

This section is divided into two parts. The first part discusses the relevant literature related to identifying important nodes in the social network, and the second part covers the metrics available for measuring network resiliency.

A. Identifying Important Structures

Identifying key individuals who are best connected or most influential in a social network is crucial for extracting actionable knowledge. Consequently, various methods have been proposed to identify these key nodes. While Hyperlink-Induced Topic Search (HITS) determines hubs and authorities [7], PageRank assigns a numerical weight for each node in the network [8]. Both of these approaches can be used to identify influential nodes. On the other hand, identifying the communities [9] and clusters from a social network perspective has also been widely studied. Generally, in a community, similar nodes are more clustered together than nodes that do not share commonalities. Previous researchers have also worked on a more sophisticated approach where their focus shifted from identifying the influential nodes or communities to detecting smaller key sets of players who maximized the information diffusion. The authors in [4] devised a methodology where they identified focal patterns leveraging the Louvain method that gave them more relevant information about the network than obtained from the influential nodes [9]. When applying this method to large biological networks, they found more prominent, smaller, and relevant structures in protein-protein interaction networks [10]. An online analysis and visualization-based tool has also been built for the ease of analyzing these small and pertinent focused structures [11]. Since this method could not extract structures with lower connection density, researchers extended their approach by combining highly connected candidate focal structures based on similarity values. This allowed the identification of both cliquish and small sparse, yet connected, structures [12]. An advanced version of this approach was proposed by [13], where the authors combined user-level centrality and group-level modularity methods to create a bi-level maximization network model that overcame the shortcomings of the previously described focal structures analysis methods.

B. Network Resiliency Metrics

Network resilience, like influential node and community identification, is crucial in Social Network Analysis (SNA), denoting a network's ability to withstand disruptions while maintaining core functions. The study by Bertoni et al. [14] employs social network analysis to identify key contributors to resilience in an intensive care unit, integrating SNA-derived indicators with non-network attributes, whereas another research comprehensively reviews resilience functions and regime shifts in complex systems across various domains through empirical observations, experimental studies, and theoretical analysis [15]. Several metrics have also been developed to quantify network resilience in the face of disruptions, such as flow robustness [16], and giant component ratio [17].

However, a key gap exists in current research. While these metrics effectively measure network resilience, they have not been extensively applied to the context of social networks like YouTube. Our work aims to bridge this gap by incorporating network resilience approaches into the analysis of social networks, offering a more comprehensive understanding of their ability to adapt and function under various stresses.

III. Data Collection

The data for this study was collected using a specialized tool designed to collect information from YouTube through its API [18]. The collection process involved retrieving videos, comments, and channel data based on specific keywords. These keywords were selected through a thoughtful process that involved reviewing commonly used terms and phrases relevant to discussions in the Indo-Pacific region. While no formal methodology was employed, the selection was guided by careful consideration of the linguistic and cultural context to ensure the keywords captured a broad range of relevant topics. Examples of keywords used include "Komunis Cina | China pengaruh Indonesia", "Muhammadiyah Cina | China | Tiongkok | Tionghoa", "Kejam Uighur | Uyghur", and "Muslim Brother | Indonesia Uighur | Uyghur". The final dataset comprised 35 YouTube channels, 308,890 videos, 726,078 commenters, and 1,536,284 comments.

IV. Methodology

This section outlines the methods used in our study. First, it details the creation of a co-commenter network and introduces focal structure analysis alongside the problem statement. Finally, it concludes by explaining various network resiliency metrics.

A. Co-commenter Network Creation

The analysis started with creating co-commenter networks for each YouTube channel. These networks connect commenters who have commented on the same video across one or more channels, as described in [19]. The edges between commenters are weighted based on the number of shared videos they have commented on. Only commenters who have engaged with at least 5 videos are included in the network, as this threshold ensures the analysis focuses on active and consistent users, minimizing noise from sporadic commenters, as shown in Figure 1.

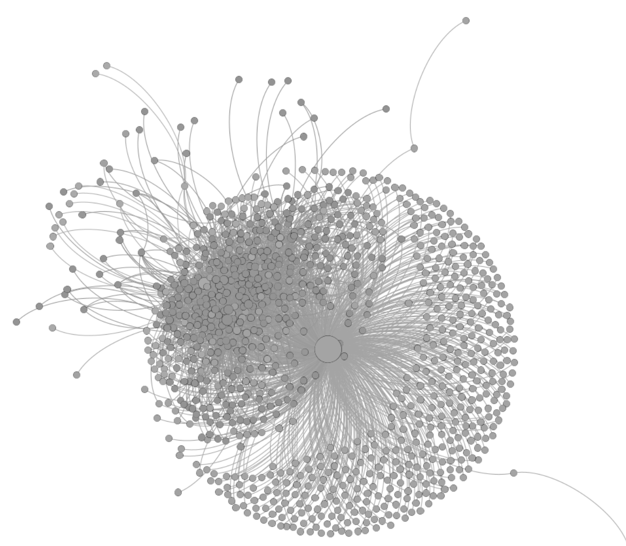

Figure 1. A YouTube co-commenter network where nodes represent commenters, and edges indicate shared commenting on the same video for the channel with ID "UCfWNZIJkm268rLO_yeRlcww".

B. Focal Structure Analysis

Focal Structure Analysis (FSA) is a social network analysis method that aims to find key sets of individuals rather than a set of key individuals within a social network. FSA aims to extract minimal influential groups in a network, thereby enhancing the knowledge discovery process. The earliest version of FSA (i.e., FSA 1.0) utilizes global and local interconnectedness-based algorithms to identify focal patterns [4]. After partitioning the network into focal structures, FSA 1.0 stitches interconnected structures using Jaccard's Coefficient [12]. FSA 1.0 groups nodes with similar clustering coefficients into focal structures, collectively identifying core-focused groups distinct from traditional community detection methods [9]. However, the current version of FSA 2.0 employs a bi-level maximization network model to identify authoritative individuals and cohesive communities within the network [13]. This analysis identifies key sets of influential commenters by leveraging degree centrality and clustering coefficient methods at the commenter level, and spectral modularity at the group level. Removing these focused core groups from the network may disrupt information flow or break down important connections, potentially compromising the overall effectiveness of its structure. Figures 2 and 3 illustrate the key focal structures identified through FSA 1.0 and FSA 2.0, respectively.

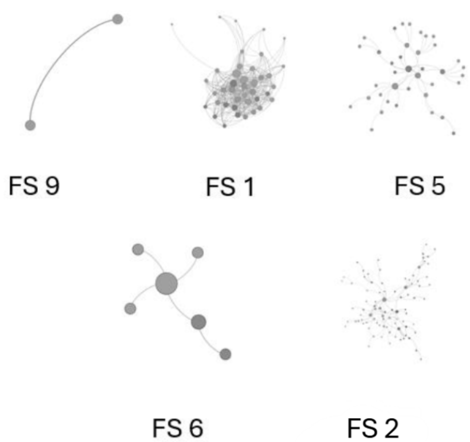

Figure 2. Several prominent Focal Structures (FS) detected using Focal Structure Analysis 1.0 from the channel with ID `"UCfWNZIJkm268rLO_yeRlcww"`.

C. Problem Statement

A Focal Structure (FS) is a key set of individuals who may be responsible for organizing information diffusion. A focal structure contains a set of vertices (at least two) and edge(s). These individuals from the focused core groups may not be the most influential on their own but by interacting together form a compelling power. Consider a social network $G = (V, E)$, where V is the set of vertices and E is the set of edges,

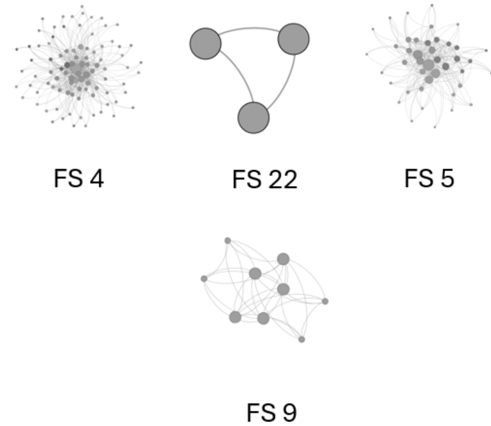

Figure 3. Notable Focal Structures (FS) uncovered through Focal Structure Analysis 2.0 from the channel with ID `"UCfWNZIJkm268rLO_yeRlcww"`.

where a focal structure can formally be defined as follows: Focal structures in G are defined by $F = \{G'\}$, where $G' = (V', E')$ and $V' \subseteq V$ and $E' \subseteq E$. For all i and j, $i \neq j$, $G_i \in F$ and $G_j \in F$, such that no two focal structures can subsume each other, or $G_i \not\subset G_j$ and $G_j \not\subset G_i$.

D. Network Resiliency Metrics

This section describes metrics used to quantify network resilience.

*1) **Flow Robustness**:* Flow robustness serves as an imperative graph metric, quantifying the resilience of a network by evaluating the proportion of reliable flows against the total flow count [16]. A flow is called reliable if it maintains at least one uninterrupted path despite potential link or node failures. It offers insight into the network's capacity to sustain communication between nodes following the removal of nodes. Flow robustness values range between 0 and 1, with 1 denoting seamless communication across all nodes and 0 indicating a lack of inter-nodal communication, indicative of a network devoid of connections. The flow robustness (FR) of a graph $G(V, E)$ is computed using:

$$FR(G) = \frac{\sum_{i=1}^{n} |C_i|(|C_i| - 1)}{|n|(|n| - 1)}, \qquad 0 \leq FR \leq 1 \quad (1)$$

*2) **Giant Component Ratio**:* The Giant Component Ratio (GCR) is a key metric in network resilience analysis, measuring the ratio of nodes within the Largest Connected Component (LCC) to the total number of nodes in the network. It is computed using:

$$GCR(G) = \frac{N_{LCC}}{N}, \qquad 0 < GCR < 1 \qquad (2)$$

where the N_{LCC} represents the number of nodes in the largest connected component and N denotes the total number of nodes in the network. This metric also serves as a critical indicator of a network's ability to maintain structural cohesion and connectivity upon the removal of focused core groups.

 ISBN: 978-1-68558-238-8

*3) **Isolated Nodes and Cluster Analysis**:* The impact of the commenter's removal from the communication network will also be evaluated through two metrics. While isolated node count measures network fragmentation that may hinder information flow, cluster analysis is performed to identify potential community fracturing and its impact on network cohesion and dynamics.

E. Spearman's rank correlation coefficient

We use Spearman's rank correlation coefficient [20], implemented to evaluate the monotonic relationship between variables without assuming linearity. The method ranks the data, assigns tied values their average rank, and computes the correlation coefficient ρ as:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(R(X_i) - R(Y_i))^2}{n(n^2-1)},$$

where $R(X_i)$ and $R(Y_i)$ are the ranks of observations in X and Y, d_i is the rank difference, and n is the number of pairs. In our analysis, the correlation between flow robustness and the giant component ratio was found to be 0.92, indicating a strong positive monotonic relationship. Given this high correlation, we selected flow robustness as the primary metric to assess the impact of nodes in each focal structure while also focusing on how the removal of a specific focal structure affects network resiliency.

V. RESULTS

This section evaluates the impact and resiliency of focal structures identified by FSA 1.0 and 2.0. We assess the influence of these structures on key metrics like flow robustness, giant component ratio and compare their performance against standard methods, such as PageRank and Louvain community detection. While both FSA 1.0 and FSA 2.0 produce key focal structures, the impact of the focal structures identified by FSA 2.0 is more prominent. Additionally, FSA 2.0 generates a greater number of such impactful structures compared to FSA 1.0.

A. Node Impact Assessment

Our study assesses the impact of each focal structure through the nodes associated with it. At first, we employed the provisional removal of each focal structure from the network and observed changes in Flow Robustness (FR) and the Giant Component Ratio (GCR). Given the strong correlation (0.92) between GCR and FR, we chose to focus on the flow robustness metric to simplify the analysis. After that, we calculated the impact score by dividing the complement of FR by the number of nodes in each focal structure. This approach allowed us to rank focal structures based on the impact of nodes within each focal structure.

Our findings reveal a noteworthy outcome where the focal structures identified by FSA 2.0 demonstrated a higher overall impact than those identified by FSA 1.0, indicating the more significant influence of nodes within these structures. Overall, these differences highlight the varying capabilities of FSA 1.0 and FSA 2.0 in revealing critical focal structures, with FSA 2.0 offering a more extensive and impactful identification of key groups within the network, as demonstrated in Figures 4 and 5.

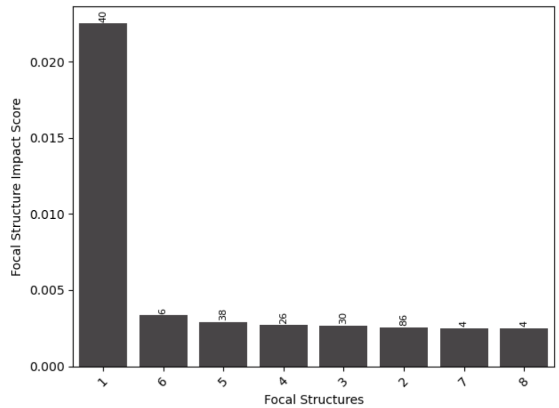

Figure 4. The calculated impact scores for focal structures identified by FSA 1.0 show the relative influence of each structure in maintaining network robustness.

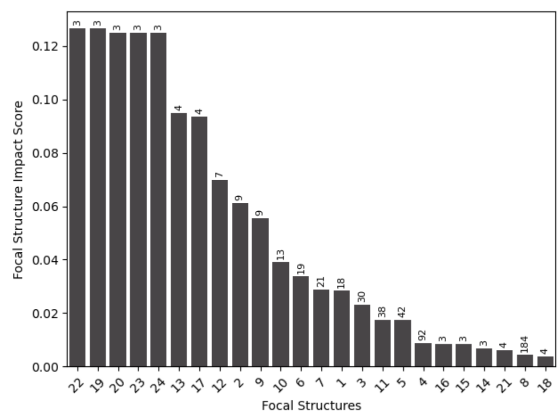

Figure 5. The impact scores of focal structures identified by FSA 2.0 demonstrate the significant role of individual nodes in affecting network resilience.

B. Network Resiliency Assessment

In this study, we also assessed the impact of focal structures identified by FSA 2.0 on network resiliency using flow robustness and Giant Component Ratio (GCR). For comparison, we evaluated the resiliency of structures detected by PageRank, the Louvain community detection algorithm, and FSA 2.0. FSA 2.0 identified 24 focal structures for our YouTube co-commenter network, compared to 7 detected by the Louvain community detection algorithm, with the top 24 influential nodes from PageRank also included for visualization purposes. Our analysis revealed that while larger community-based structures, such as those identified by the Louvain algorithm, contained more nodes, they did not exhibit the same impact on the network as the focal structures identified by FSA 2.0. The focal structures from FSA 2.0 consistently outperformed the community structures regarding flow robustness and GCR. Figure 6 reveals that focal structures from FSA 2.0 consistently result in a greater reduction in flow robustness, highlighting their critical role in maintaining information flow, whereas community and PageRank nodes exhibit comparatively lower impact. Furthermore, Figure 7 shows a pronounced decrease in GCR upon the removal of FSA 2.0 focal structures, underscoring their significant influence in sustaining the largest connected component, while community and PageRank nodes exhibit less disruptive effects. This finding underscores the unique contribution of focal structures built on individual and group-based node features. In addition to that, focal structures play a crucial role in bridging communities and maintaining overall network connectivity regardless of their size.

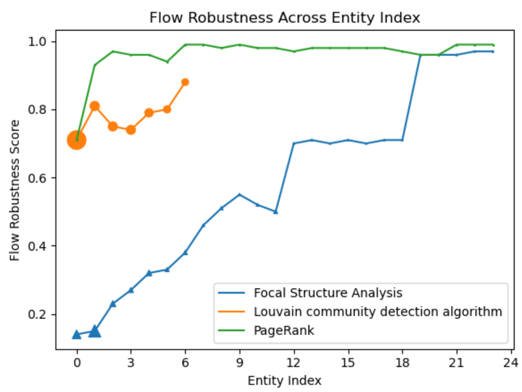

Figure 6. Comparison of the impact on network flow robustness when removing key structures identified by FSA 2.0, Louvain community detection, and PageRank.

The network's modularity increased when we removed the impactful focal structures that FSA 2.0 had found. This, in turn, indicates that these structures are essential to maintaining the information flow across communities. On the other hand, the removal of the larger community structures did not have the same impact. Their function as crucial gatekeepers in the spread of information is further highlighted by the network fragmentation brought about by the removal of smaller, well-positioned comments from FSA 2.0 networks.

This comparison demonstrates how much better FSA 2.0 is at locating critical structures that have a big impact on network resilience. The focal structures identified by FSA 2.0 continuously shown noticeable influence on the network, showing their crucial role in preserving information flow and network cohesion, even if community-based structures had a larger number of nodes.

Lastly, when provisionally removing focal structures identified by FSA 2.0, it caused considerable fragmentation within the network, leading to the isolation of nodes from the overall network and the formation of numerous clusters. For instance, the removal of one focal structure resulted in 611 clusters and 605 isolated nodes, while even a focal structure containing only 3 nodes was able to isolate 431 nodes, as shown in Table 1. These findings underscore the imperative influence of focal

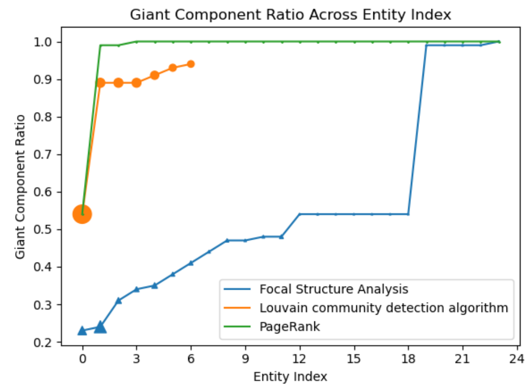

Figure 7. Evaluation of network resilience through changes in the Giant Component Ratio (GCR) following the removal of nodes detected by FSA 2.0, Louvain community detection, and PageRank.

TABLE I. IMPACT OF REMOVING PROMINENT FOCAL STRUCTURES IDENTIFIED BY FSA 2.0 ON NETWORK FRAGMENTATION.

Focal Structure	Nodes	No. of Clusters	Isolated Nodes
5	42	611	605
9	9	486	483
22	3	434	431

structures in the network, revealing how their removal can disproportionately disrupt connectivity and lead to significant fragmentation, even when the focal structure itself is relatively small. As a result, focal structures play a pivotal role in preserving network connectivity and highlight their significant impact on maintaining overall network cohesion.

VI. Conclusion and Future Work

In this study, we first obtained the focal structures of YouTube co-commenter's network by leveraging two distinct versions of focal structures analysis, FSA 1.0 and FSA 2.0. Furthermore, through various network resiliency metrics, we delved deeper to assess how these focal structures were crucial to the overall success of information dissemination for the defined networks. By examining the flow robustness and giant component ratio, we demonstrated that the focal structures detected by FSA 2.0 exhibit a significantly higher impact on the network compared to those identified by both FSA 1.0 and traditional community detection algorithms. Despite the larger size of community-based structures, they failed to match the influence of smaller, strategically positioned focal structures identified by FSA 2.0. Our evaluation not only justified that removing some of the focal structures from the network made it more sparse, fragmented, and less cohesive but also that the information flow of the co-commenter network was disrupted heavily, which means that these focal structures were acting as a bridge between other commenters of the communication network.

In future, to advance our understanding of commenter networks and focal commenter groups, our research should

 ISBN: 978-1-68558-238-8

utilize the contextual representation of commenter networks by incorporating content, engagement, and other attributes. Utilizing contextualized focal structure analysis could thus help enhance the comprehensive discovery and interpretability of focal commenter structures.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, 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. 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. I. Gurung, M. M. I. Bhuiyan, A. Al-Taweel, and N. Agarwal, "Decoding youtube's recommendation system: A comparative study of metadata and gpt-4 extracted narratives," in *Companion Proceedings of the ACM on Web Conference 2024*, 2024, pp. 1468–1472.

[2] M. I. Gurung, N. Agarwal, and M. M. I. Bhuiyan, "How does semiotics influence social media engagement in information campaigns?," 2025.

[3] S. Shajari, M. Alassad, and N. Agarwal, "Characterizing suspicious commenter behaviors," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 2023, pp. 631–635. DOI: 10.1145/3625007.3627309.

[4] F. Sen, R. T. Wigand, N. Agarwal, D. Mahata, and H. Bisgin, "Identifying focal patterns in social networks," in *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, IEEE, 2012, pp. 105–108.

[5] 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 *Modeling and Simulation of Social-Behavioral Phenomena in Creative Societies: First International EURO Mini Conference, MSBC 2019, Vilnius, Lithuania, September 18–20, 2019, Proceedings 1*, Springer, 2019, pp. 41–54.

[6] S. Shajari, R. Amure, and N. Agarwal, "Analyzing anomalous engagement and commenter behavior on youtube," *AMCIS 2024 Proceedings. 6*, 2024.

[7] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[8] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, P10008, 2008.

[10] F. Şen, R. T. Wigand, N. Agarwal, M. Mete, and R. Kasprzyk, "Focal structure analysis in large biological networks," in *IPCBEE, ser. 1, vol 70, IACSIT Press.*, 2014.

[11] F. Sen, N. Nagisetty, T. Viangteeravat, and N. Agarwal, "An online platform for focal structures analysis-analyzing smaller and more pertinent groups using a web tool," in *2015 AAAI Spring Symposium Series*, 2015.

[12] F. Şen, R. Wigand, N. Agarwal, S. Tokdemir, and R. Kasprzyk, "Focal structures analysis: Identifying influential sets of individuals in a social network," *Social Network Analysis and Mining*, vol. 6, pp. 1–22, 2016.

[13] 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," *Central European Journal of Operations Research*, vol. 30, no. 1, pp. 367–394, 2022.

[14] V. B. Bertoni, T. A. Saurin, and F. S. Fogliatto, "How to identify key players that contribute to resilient performance: A social network analysis perspective," *Safety science*, vol. 148, p. 105 648, 2022.

[15] X. Liu *et al.*, "Network resilience," *Physics Reports*, vol. 971, pp. 1–108, 2022.

[16] M. J. Alenazi and J. P. Sterbenz, "Comprehensive comparison and accuracy of graph metrics in predicting network resilience," in *2015 11th international conference on the design of reliable communication networks (DRCN)*, IEEE, 2015, pp. 157–164.

[17] B. Bollobás and B. Bollobás, *Random graphs*. Springer, 1998.

[18] O. Adeliyi, I. Solaiman, S. Shajari, U. Onyepunuka, and N. Agarwal, "Detecting and characterizing inorganic user engagement on youtube," in *Workshop Proceedings of the 18th International AAAI Conference on Web and Social Media*, 2024. DOI: 10.36190/2024.01.

[19] S. Shajari, N. Agarwal, and M. Alassad, "Commenter behavior characterization on youtube channels," in *Proceedings of the eKNOW International Conference on Information, Process, and Knowledge Management, Venice Italy*, 2023, pp. 59–64.

[20] C. Spearman, "The proof and measurement of association between two things," *The American journal of psychology*, vol. 100, no. 3/4, pp. 441–471, 1987.

 ISBN: 978-1-68558-238-8