Finding Fake News Key Coordinators in Complex Social Networks Using Bi-Level Decomposition Optimization Method

Mustafa Alassad, Muhammad Nihal Hussain, and *Nitin Agarwal*[*]
[*]Endowed Chair and Distinguished Professor
Collaboratorium for Social Media and Online Behavioral Studies (COSMOS)
University of Arkansas – Little Rock
Email: nxagarwal@ualr.edu

Outline

- Introduction
- Problem Definition
- Methodology
- Results & Analysis
- Conclusion & Future Research

Acknowledgments

This research is funded in part by the

- U.S. National Science Foundation (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, N00014-17-1-2605, N00014-17- 1-2675, N00014-19-1-2336, N68335-19-C-0359),
- U.S. Air Force Research Lab,
- U.S. Army Research Office (W911NF-16-1-0189),
- U.S. Defense Advanced Research Projects Agency - DARPA (W31P4Q-17-C-0059),
- Arkansas Research Alliance,
- The Jerry L. Maulden/Entergy Foundation at the University of Arkansas, Little Rock.

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.

Introduction

- Social Media is an important platform for people to raise voice.
- Social Media is used in organizing major protests campaigns.
- Social Media can be used to mobilize crowds.
- Advanced social networks analysis methods are needed to study such complex behaviors.
- Lack of research in identifying higher level coordinating groups by regular community detection algorithms and node-level influence assessment approaches.

Definition

Focal Structure Analysis

It is a method to identify coordinating structures that are usually undetected by community detection algorithms.

It explores the smallest possible sets of individuals that can influence maximum number of users in social networks.

These sets of individuals, when coordinating together, maximize information diffusing, influence operations, mobilizing crowds.

Background

Śen et al. (2016) – FSA V1.0

- Applied a greedy algorithm;
- Investigated the Egyptian Revolution (2011);
- Five coordinating groups.

Alassad et al. (2019) – FSA V2.0

- Applied a Bi-Level optimization Method;
- Investigated a YouTube Co-commenter Network;
- Found small intensively coordinating groups of commenters.

Problem Definition

1 **Node Based Detection Algorithm, Centrality Method**

Node_Level

Maximizing Node's Centrality

2 **Group Based Detection Algorithms, Modularity Method**

Network_Level

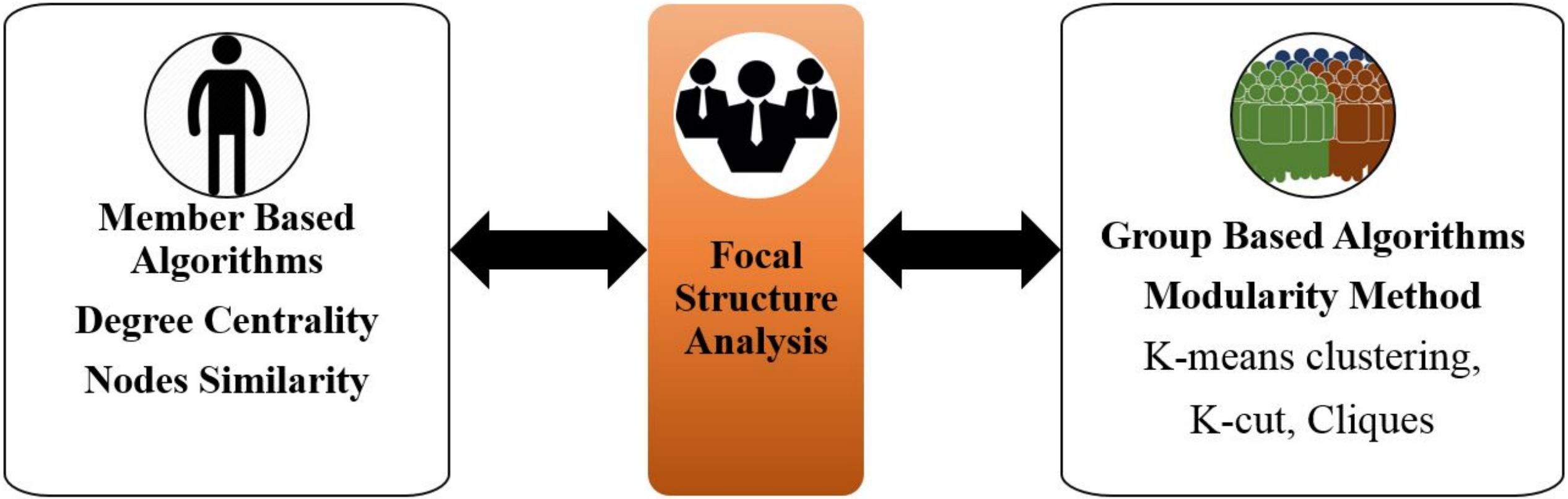
Maximizing Graph's Modularity

3 **Extract Key Focal Structure Sets, Max – Max optimization method**

Node_Level & Network_Level

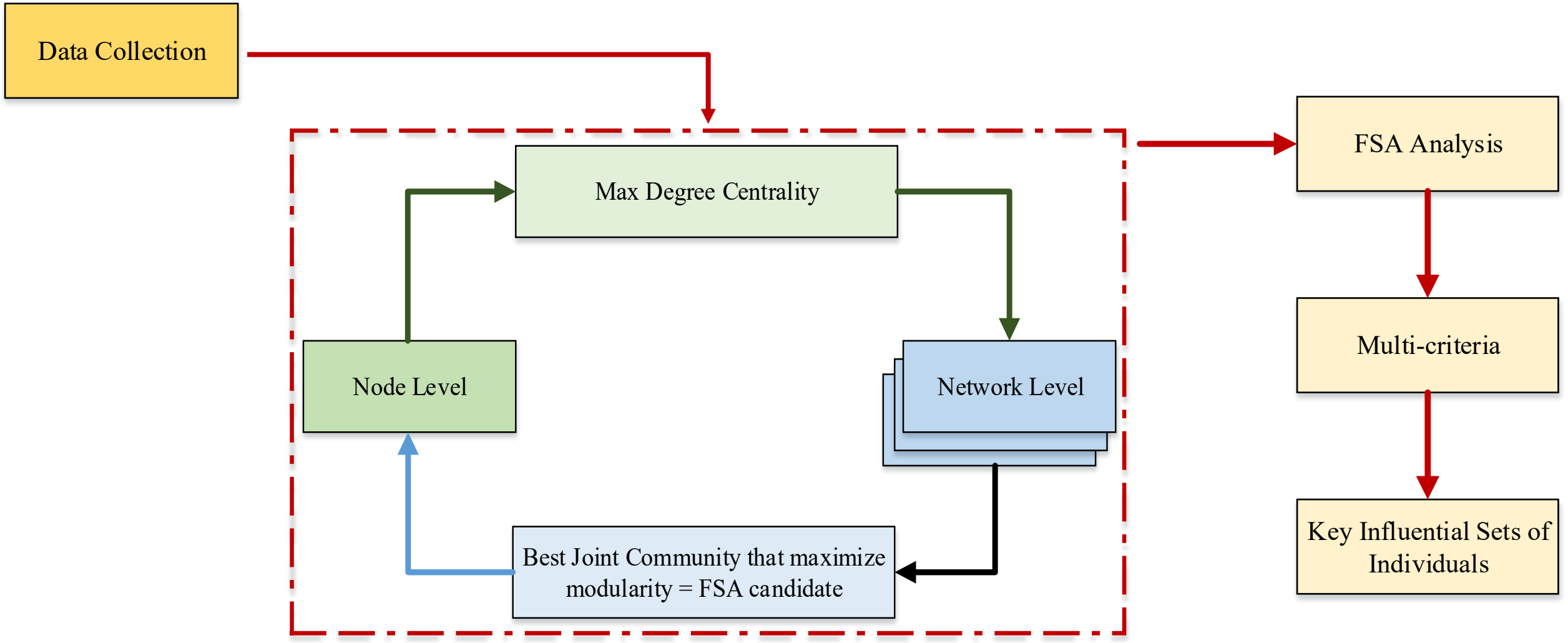
Extract Key Sets that include central nodes and maximize graph's modularity

Problem Definition



Focal Structure Analysis

Research Methodology



Bi-level Optimization

Node-level optimization

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

Subject to

$$\delta_i = \{d\vec{c}_1 \preccurlyeq d\vec{c}_2 \preccurlyeq d\vec{c}_3 \preccurlyeq \cdots \preccurlyeq d\vec{c}_i\} - \overline{dc_j^Q} \quad \forall i,j \quad (2)$$

$$d_i^c = \sum_j m_{ij} \quad \forall i \quad (3)$$

$$d_i^c \geq 2 \quad \forall i \quad (4)$$

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

$$D_G^L < \hat{d}_i^c \leq D_G^U \quad \forall i \quad (6)$$

$$a_i^c = \frac{(\#\ of\ Triangles) \times 3}{\#\ of\ Connected\ Triples\ of\ Nodes} \quad \forall i,j,z \quad (7)$$

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

$$AC_G^L < a_i^c \leq AC_G^U \quad \forall i \quad (9)$$

$$\overrightarrow{C_v} = \{\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{c_3}, \dots, \overrightarrow{c_i}\} - \overrightarrow{c_j^Q} \quad \forall i,j \quad (10)$$

$$\overrightarrow{c\delta_i} = \overrightarrow{C_{\delta_i}} \quad \forall i \quad (11)$$

$$F = \left\{c_I, \overrightarrow{c_j^Q}, \overrightarrow{c_{j+1}^Q}, \dots, \overrightarrow{c_\kappa^Q}\right\} \quad \forall j,\kappa \quad (12)$$

Network-level optimization

$$max \sum_{j=1}^{n} \varrho_j^M \quad (13)$$

Subject to

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

$$\delta_j = \{\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{c_3}, \dots, \overrightarrow{c_n}\} - \overrightarrow{c\delta_i} \quad \forall j \quad (15)$$

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

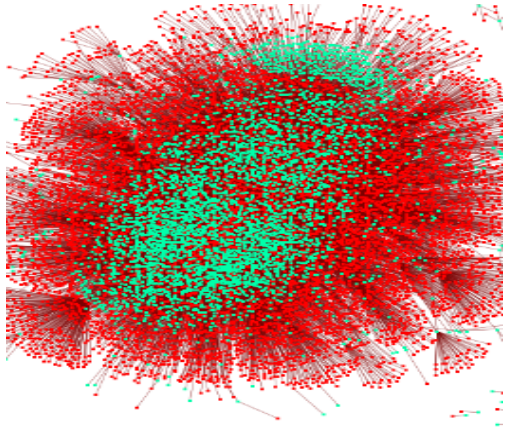
$$\xi_j = \{\overrightarrow{c\delta_i} \cup \delta_j | \overline{\delta_i} \neq \delta_j\} \quad \forall j \quad (17)$$

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

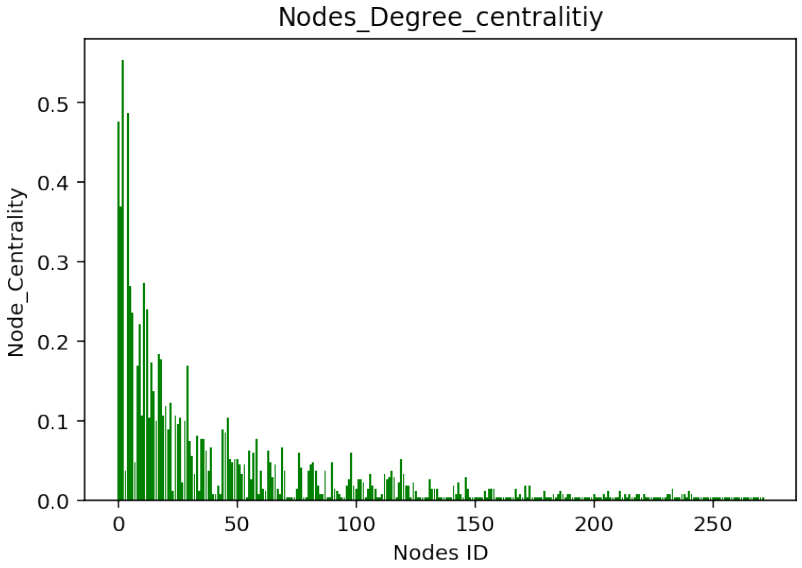
$$\varrho^l \leq \varrho_j \leq \varrho^U \quad \forall j \quad (19)$$

Data Collection

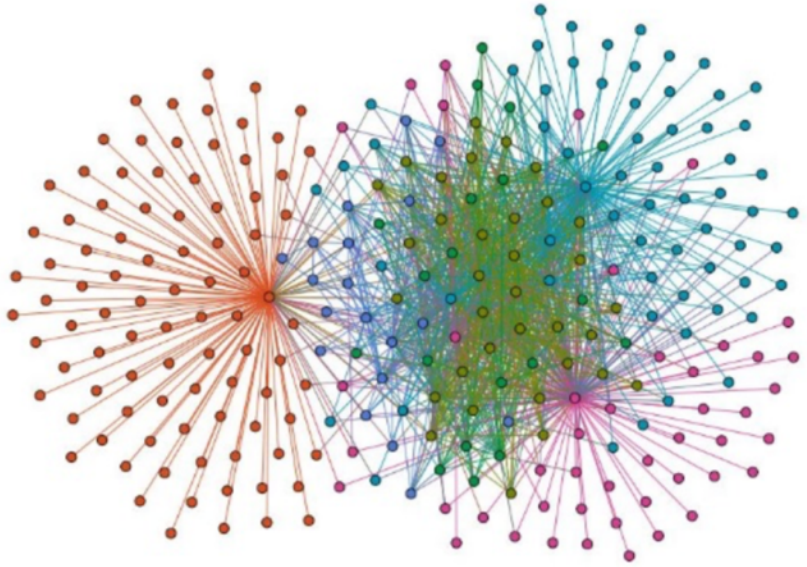
- YouTube Channel “Hot News” spreading fake news related to the conflict in the South China Sea.
- Joined August 13, 2016
- 15 millions views
- 5,095 videos (green colors)
- 8,477 commenters (red color)
- 47,265 comments
- Network is an undirected network of YouTube commenters.
- More than 1 million edges



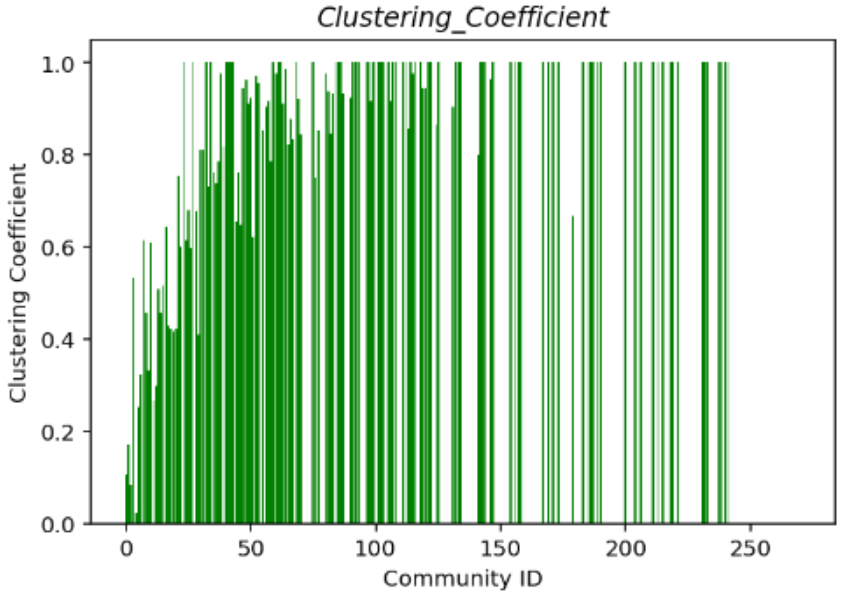
Results



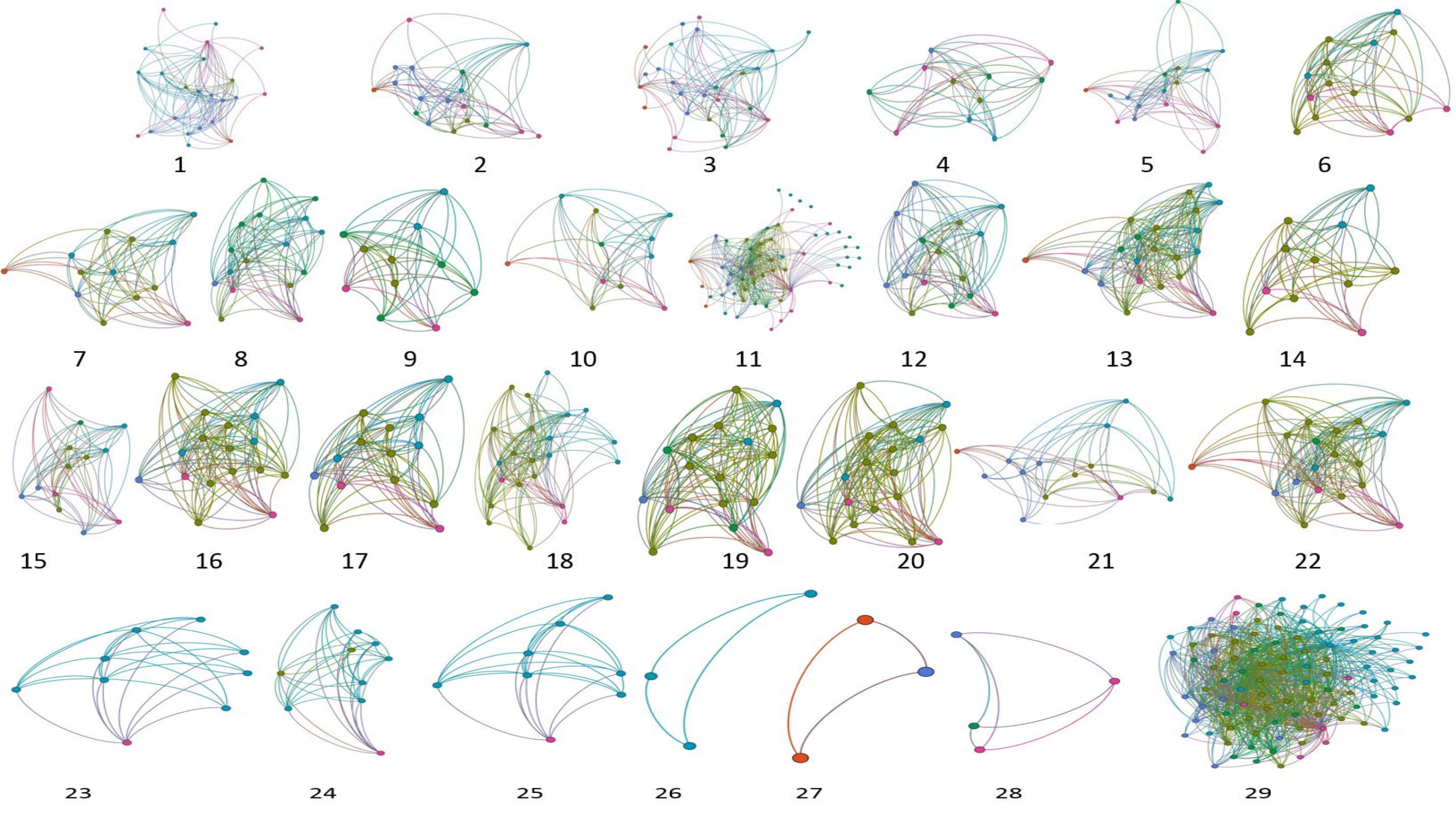
Degree Centrality Measures



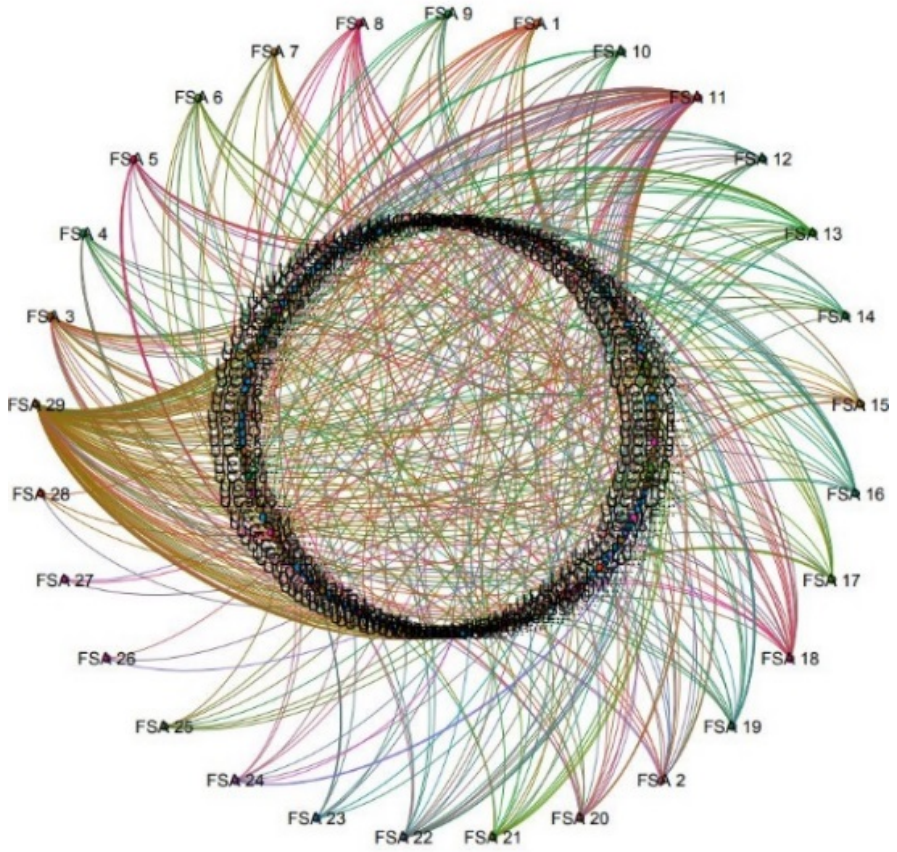
Commenter Network of >= 10 Shared Videos



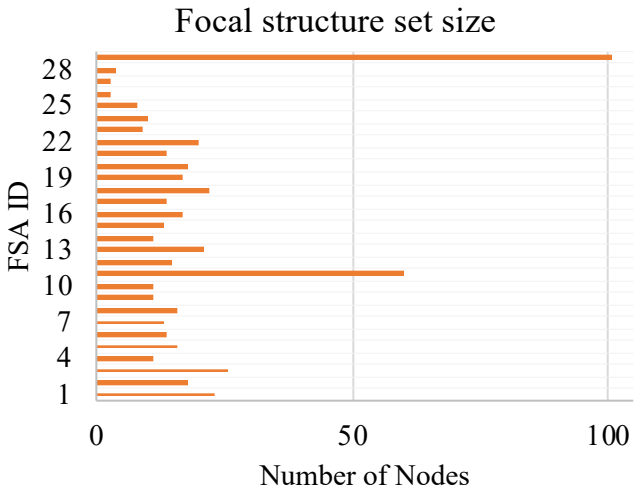
Avg. Clustering Coefficient Values


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

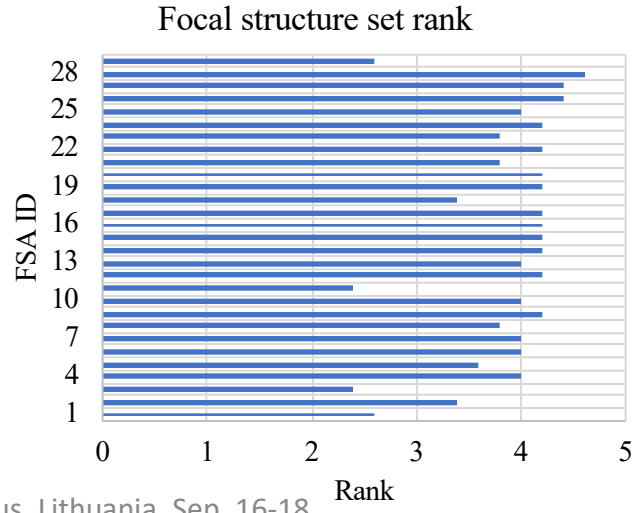
Analysis of Focal Structure Sets



Focal structure sets identified from the YouTube dataset

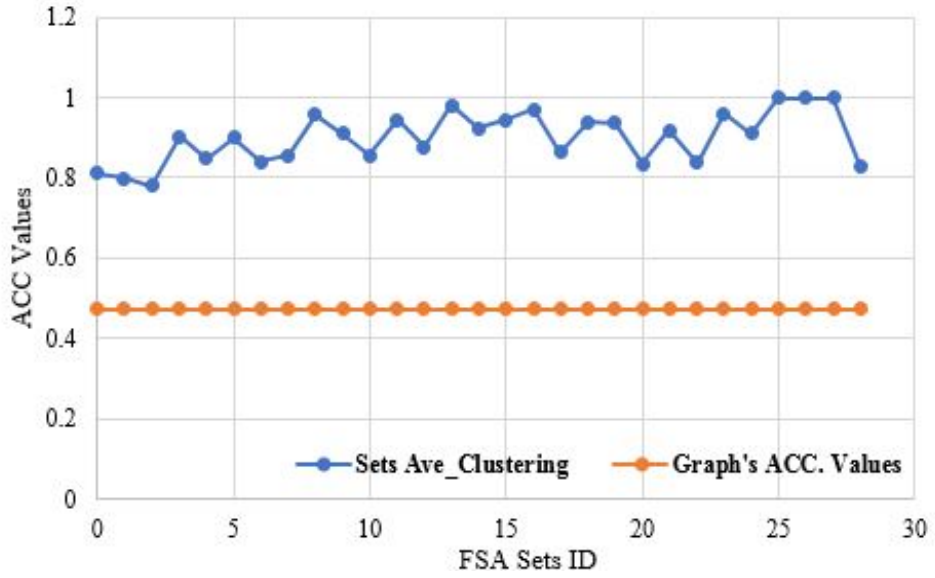


Identified focal structure sets size.

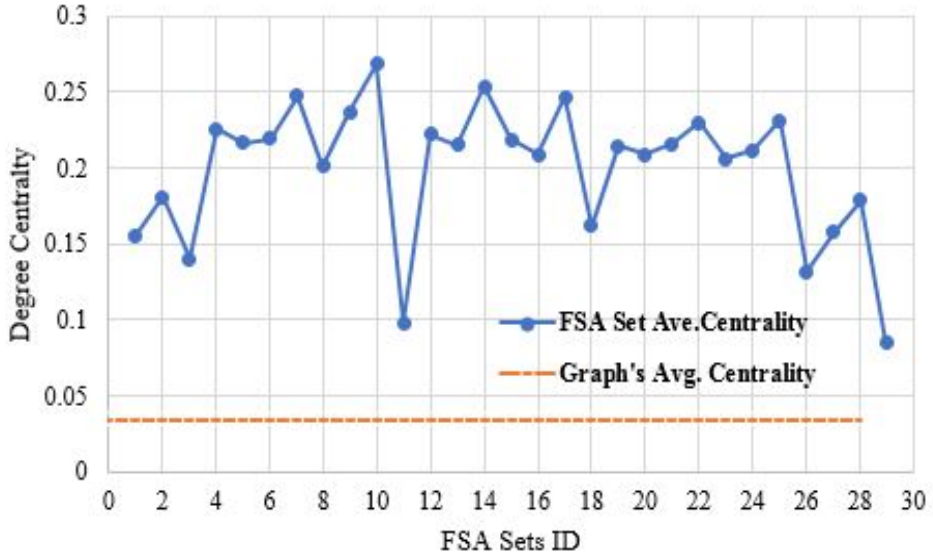


Most of the identified focal structure sets have high rank.

Analysis of Focal Structure Sets



Average Clustering Coefficient (ACC) of Focal structure sets, vs. graph.

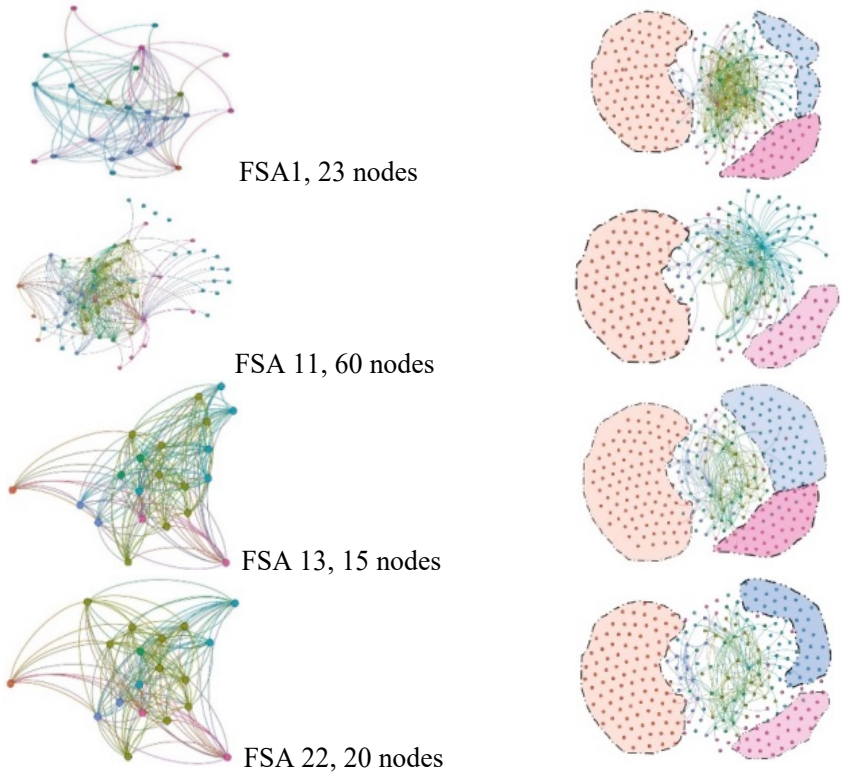


Average Degree Centrality of Focal structure sets vs. graph value.

Key Set Performance

FSA Key Sets have the control & power to spread fake news because,

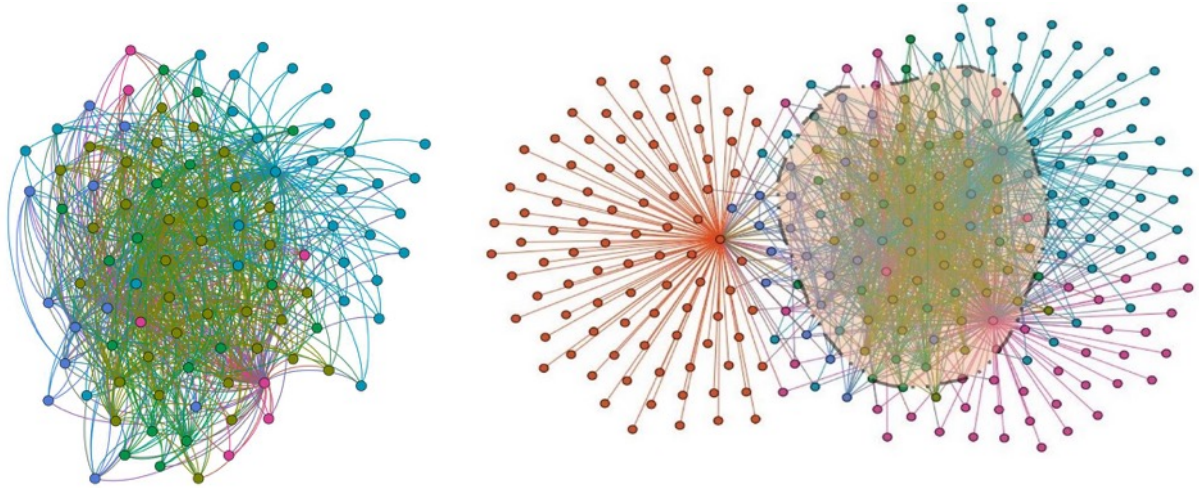
- FSA Key sets have high Interactions with other members;
- FSA Key sets have high intra-group connections;
- FSA Key sets have high centrality nodes;
- FSA Key sets could sparse the graph easily.



Demonstrating the influence of key sets or focal structures of commenters for a YouTube Channel.

Salient Features of the Model

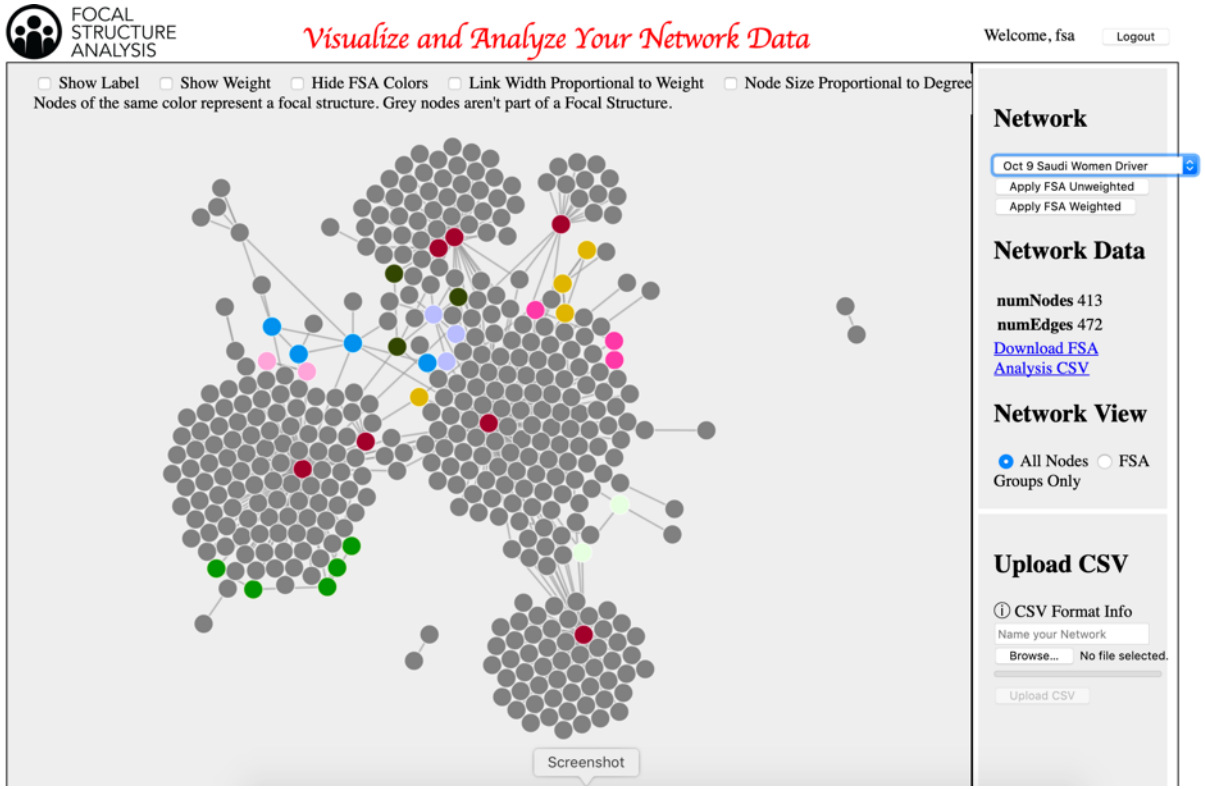
- Unsupervised, non-parametric model;
- No probabilities, and estimations within the model;
- Linear decomposed model to maximize individual's centrality and graph modularity;
- Focus on central nodes' connections;
- Identifies hidden communities; and
- Simplify complex network analysis;



Modularity method did not identify structure 29 (left side), one of the significant focal structures identified by the FSA algorithm implemented in this paper.

Focal Structure Analysis Software

- Web-based application
- No installation needed
- Upload network in csv format
- Identifies focal structures and extract/export



http://blogtrackers.host.ualr.edu/fsa/

Summary

- Applying a bi-level decomposition optimization to maximize the degree centrality and graph's modularity.
- Identified small key influential sets of fake news spreaders.
- These small sets can include influential members acting in different groups.
- Non-influential members become influential within focal structure sets.
- Small sets acting like real world organization.
- They have high interactions with other groups and, high intra-group connections.

Future Research Plans

- Leverage social science theories of collective action to measure the power of individual focal structure;
- Investigate FSA application in other social media networks and other domains (e.g., email networks, financial networks, organization networks, power grids); and
- Implement dynamic and predictive approaches.

Thank You

Email: nxagarwal@ualr.edu