**Solid State Technology**
**Volume: 63 Issue: 6**
**Publication Year: 2020**

# Online social Network Trend Discovery Using Frequent Subgraph (FSgM) Mining

**[1]S. Ajay Kumar, [2]P.V. Ramana Murthy, [3]Pattola Srinivas, [4] P.Andrews Hima Kiran**

*[1]Research Scholar, KLEF Vaddeswaram &Assoicate Professor, CSE Department,Malla Reddy Engineering College (A),Hyderabad, Telangana, India*
*[2]Research Scholar, Rayalaseema University & Assoicate Professor, CSE Department,Malla Reddy Engineering College (A),Hyderabad, Telangana, India*
*[3&4]Assoicate Professor, CSE Department,Malla Reddy Engineering College (A),Hyderabad, Telangana, India*

***Abstract***- A social network (SN) is an online platform where people use to contact dynamically with various social networks and establish social relationship with other people and share similar personal, career interests, activities, and real-life connections.An OSN is a dedicated website or application used to communicate with each other. In today's internet research world due to the advancements in computer hardware and efficient computing processing power, construction of large-scale graph data mining has become a plus point to graph mining researchers. So Graph mining has become one of the most important disciplines in the research area of data mining.A social network contains a huge amount of unstructured data resulting to complex challenging problem to many researchers. A Frequent Sub graph mining (FSgM) techniques is been used to identify the frequent pattern trends existing in the network for networks analysis. A Frequent Sub graph mining (FSgM). is playing a very prominent role in the core graph operation with its domain areas like social networking analysis, web data mining bioinformatics, graph data management, knowledge exploration and security. Frequent subgraph mining have extremely high computational complexity. Finding subgraph patterns for frequently reappearing social network graph dataset is important to identifying the interpretable structural properties of complex networks to study the trend and also for social balance and status findings. Many researchers have proposed many graph mining algorithms, but we observed a few are working for capturing the important element of Social networks, so this became trends to discovery of frequent pattern mining. In this paper we introduces a novel FSgM approach, called gSpan (Graph Based Substructure Pattern Mining) to identify the frequently occurring pattern trends in the social network data. We considered Facebook social media data for identifying the frequent pattern trend and performed analysis.

**Keywords** *Graph mining, frequent pattern trends, frequent subgraph mining, Social network analysis, graph data management, bioinformatics,*

## 1. INTRODUCTION

We all have moved from the days of MySpace to a social media era dominated by Facebook, Twitter and other social mobile applications. Accessing to the Social networks has become a trend and a meeting point for today's internet savvy audience. Today we see Major portion of younger generation, teenagers and middle aged people, are major percentage of the total social media user population. Social media is proving open possibilities to all for direct access to clients without any third party intervention.[1] Advertising through social media is pretty cost friendly as compared to costs incurred by print, TV or other traditional media. Social media also helping the search engine to optimization and increase in rankings of any company websites. So we observe data from many forms like structure and unstructured data have been extensively growing in our day to day life. So this leads tom the researcher to involve in knowing the trends in today's social media networking works. This complex problem Graph mining has become one of the most important disciplines in the research area of data mining. Representing the today's internet world real- world problem can be done more efficiently using data mining Graph-based approaches. Graph mining has become a well established discipline within the domain of data mining.

Graph mining has been further sub-classified into following sub areas.

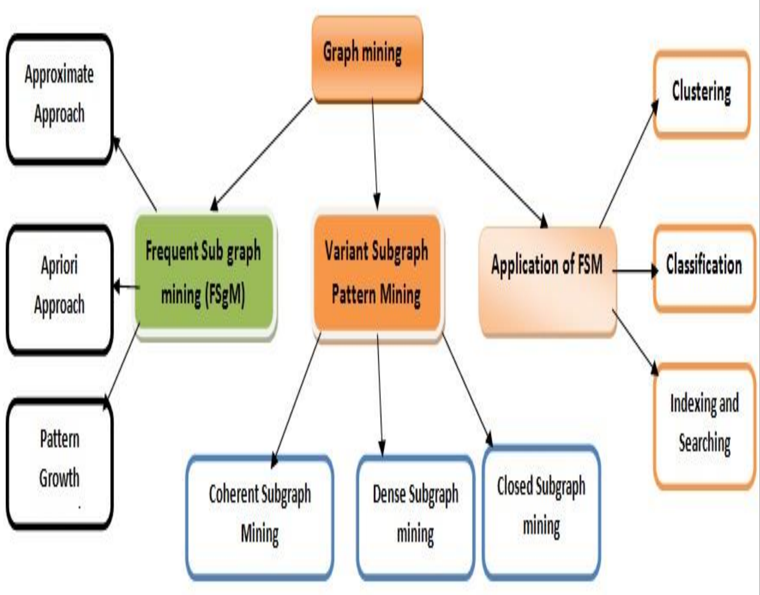



**Figure 1: Graph Mining Categories**

We considered (FSgM) data mining approach for our social networks trend discovery because it is the core of graph mining and mostly used for graph cluster analysis, characterizing of the graph sets, differentiating between different given graph groups.

**1.1 Frequent sub graphs Mining:** A (sub) graph is frequent if its support the occurring frequency in a given dataset is no less than a minimum support threshold and Support of a graph g is defined as the percentage of graphs in G which have g as subgraph[8]

Online social network (OSN) is a website/application which brings people together to communicate, share ideas and interests, or make new friends. Social networks can be used to find the interesting things on the Internet as your friends and family likely share many of your interests.

An online social network consists of some social entity so called as actors and are connected to each other by means of one of several links.

The first true social networking application is SixDegrees.com, created by Andrew Weinreich in the year 1997.W e have different kinds of online social networking platforms available today in this internet world and have to follow various approaches to discovered the trends in these social networks.

Some of the famous social networks are listed in the figure 2.

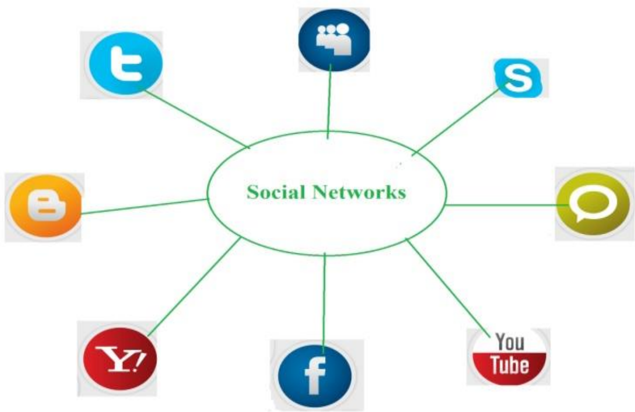



**Figure 2: Social networks**

Our different studies among these online social networks signified valuable and innovative insights proving that Facebook is the most well-liked social networking platform with highest number of users.

Currently Social network analysis is used widely used in the areas of social, behavioral sciences, economics, marketing, industrial engineering, and in military. These are the most benefiting areas of today's day to day world from the social analysis tools.

Social network analysis if found to be more useful to military , criminal networks to discover the user interest and can able to identify the potential impact in the social relations. By doing this with criminal networks we can able to identify the frequent nodes (Users) activities within the considered network, and can now skip the other nodes (user). We may be further concerned with these specific nodes to disseminate the data in the network.

Consequently, this will help the social networks to further enhance their platforms for the betterment of the users as well as for their businesses. We implemented this graph mining pattern extraction paper with 5 sections. In the next we have given the basic groundwork required for understanding the graph mining and graph terminology for representing the problem statements and solution. Section 3 briefly described the related work, which have motivated our approach. In Sect. 4, we present the details on the proposed frequent pattern trend mining FSM framework, A-RAFF. In Sect. 5, frequent pat- tern trends are discovered and evaluated from real-world online social networks. Finally, in Sect. 6, we draw conclusions with final thoughts for future work.

## 2 .ASPECTS OF GRAPH MINING

In graph mining approach we use maximum graph representation. A graph is a collection of nodes and links between nodes. A Graph supports all the aspects of a relational data mining process by representing the entities, attributes and their relationships to other entities. Graph also stores relationship information of about their connected entities. Using a graph for representing the data and the mined knowledge supports direct visualization and increases comprehensibility of the knowledge. Mining a graph data is one of the most capable approaches to extract knowledge from relational data.

**2.1Mining graph data:** It is a new scientific area. Its main goal is to present the most significant methodologies for discovering patterns in graph-structured data and subsequently to highlight some strategic application domains. Mining graph data is also known as graph-based data mining. It is the process of extraction a meaningful and useful knowledge from a graph representation of data. We use a (FSgM) approach for mining the facebook data sets to extract the frequent patterns to provide significant results[9]

**2.2 Mining frequent patterns:** A pattern is a set of items, subsequences, substructures that occurs frequently in a data set.

## 3. TERMINOLOGY USED FOR GRAPH MINING

In computer science terminology a network that consists of a finite number of users (nodes) who are interacting with each other (edges) and shares information (labels, signs, edge directions ).[ 6]

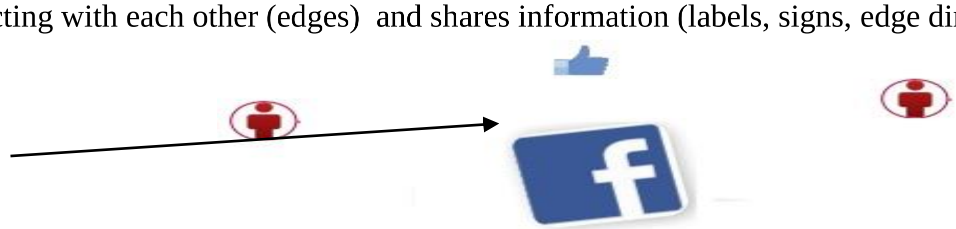


**Figure 3: simple graph notation showing two facebook users connecting**

We provide some of the basics of graph theory and its enablers.

**3.1 Definition**: **Vertex/Node**: A user of the social network (Facebook user)

**3.2 Edge/Link/Tie**: The connection between two vertices referring to their relationship or interaction.

**3.3 Graph**: A Graph is a pair of sets. G = (V,E). V is the set of vertices. E is a set of edges. E is made up of pairs of elements from V (unordered pair).

**3.4 Subgraph:** A graph G2 = ( V2, E2 ) is a subgraph of another graph G1 = ( V1, E1 ) iff V2⊑V1 and E2⊑E1 ∧ ( v1, v2 ) ∈ E2 → V1 ∈ V2 and v2 ∈ V2. G1 is called a super-graph of G2.
**3.5** A **Digraph** is also a pair of sets. D = (V,A). V is the set of vertices. A is the set of arcs. A is made up of pairs of elements from V (ordered pair)[10]

**3.6Ordered graph** : A graph *G* is an ordered pair *G* = (*V*, *E*) preserving the following conditions, *V* is a set, whose items are called vertices or nodes and *E* is a set of unordered pairs of vertices, called edges. Graph G=(V,E)

**3.7 Social network graph:** A Social Network is a labeled graph. $\sum_V$ is the label of the vertices; L is used as a labeling function:
$DG = (V, E, \sum_V, L)$

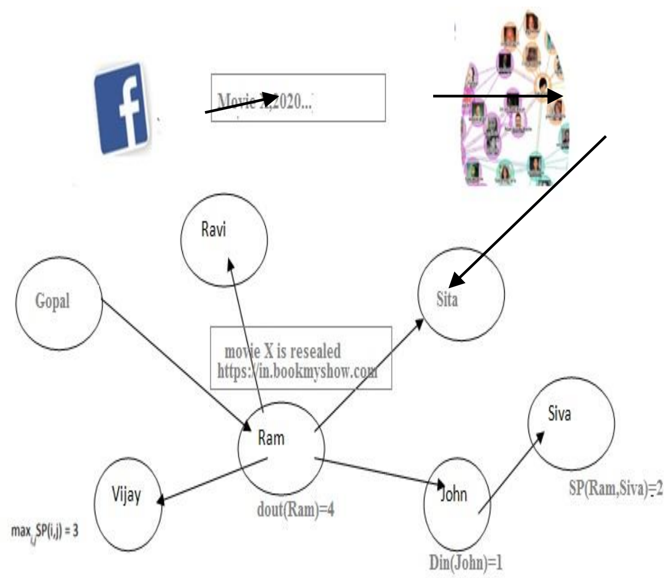



**Figure 4: Social network graph**

**3.8 Ordered Graph**: we can represent the graph order by a notation |*V*| and |*E*| denoting the size of the graph.

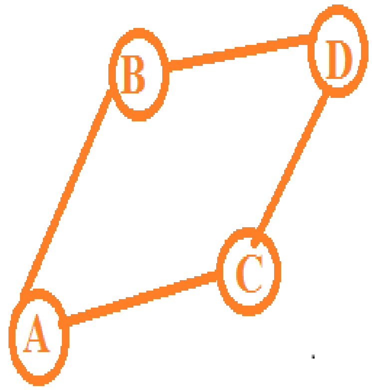



**Fig 5: Ordered graph with 4 nodes, $V = (A, B, C, D)$ & 4 edges $E = \{(A, B), (A, C), (B, D), (C, D)\}$**

**3.9. Weighted graph:** In this type of graphs a numerical values is assigned with the edges of the graph representing as weights of the graph. Weights represent the distance between two nodes, and cost of travelling from one node to other or time taken to reach from one to other node.

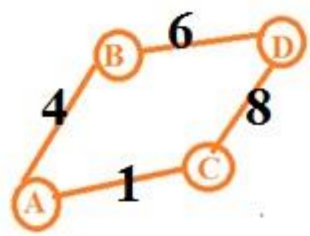



**Fig 6: Weighted Graph**

## 4. Representation of Interaction and concepts in social networks

| | |
|---|---|
| **Roles** | **Friend of, Brother/Father of, Co-worker of** |
| **Affective** | **Likes/Dislikes, Loves/Hates** |
| **Proximity** | **Knows, Shares interests** |
| **Actions/Communication** | **Talks to, Comments on, Is with** |

## 5. Interesting concepts and patterns in the network as graphs

A picture speaks a thousand words like wise a graph speaks so much more than a picture.

A visual representation of data, in the form of graphs, helps to gain actionable insights and make better data driven decisions.

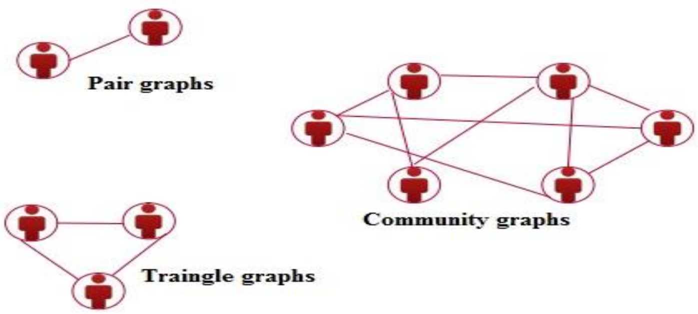



**Figure 7: Graph patterns in networks**

## 6. Graph Based Substructure Pattern Mining

In this paper we introduces a novel FSgM approach, called gSpan (Graph Based Substructure Pattern Mining) to identify the frequently occurring pattern trends in the social network data. We considered Facebook social media data for identifying the frequent pattern trend and performed analysis.

We Used gSpan to reduce the generation of duplicate graphs. gSpan is a complete frequent subgraph mining algorithm. gSpan improves the performance over straightforward apriori extensions to graphs through DFS Code. gSpan uses a depth first search (DFS) search techniques to travel through the connected nodes in the considered graph.

DFS lexicographic order and minimum DFs code forms a code form a canonical labeling system to support DFS search. This approach is used to discover all the frequent subgraphs without a candidate generation and false positive pruning. It combines the growing and checking frequent sub graphs in to one procedure thus accelerating the mining process.

### 6.1 gSpan: Graph Based Substructure Pattern Mining process:

Step 1: starting vertex is chosen randomly

Step 2 Vertices in a graph are marked so that we can tell which vertices have been visited.

Step 3: visited vertex set is expanded repeatedly until a full DFS tree is built

Step 4: Given a graph G and a DFS tree T in G , a new edge e can be added between the right most vertex and another vertex on the right most path (backward)or can introduce a new vertex and connect to a vertex on the right –most path(forward)

By using a pattern growth based frequent substructure mining we can reduce duplicate graphs[5]

### 6.2 Algorithm gSpan

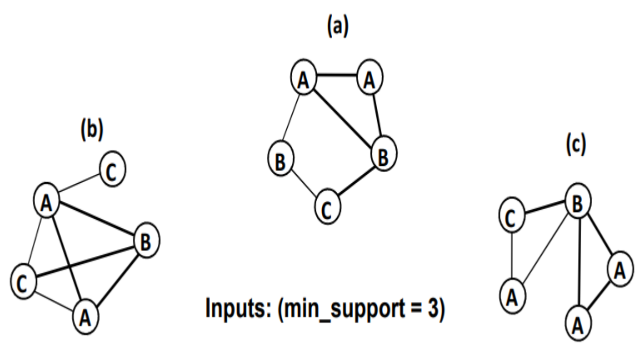
**Input**: s, a DFS code, D- a graph data set, min_support a threshold

Procedure gSpan(s,D,min_sup,S)
if s ≠ dfs(s) then return;
insert s into S;
Set C to Φ
Scan D once, find all edges e that s can be right-most extended to S $_r$ e
insert s$\diamond_r$ e into C and count its frequency;
for each frequent s$\diamond_r$ e in C do
gSpan( s$\diamond_r$ e D, min_sup,S)
return;

**output :**

S, frequent graph set
Method
S⟵Φ
Call gSpan( s,D,Miin_sup,S)

**gSpan Algorithm with minimum support =3**




With minimal support =3 we obtained a graph as shown in fig.8

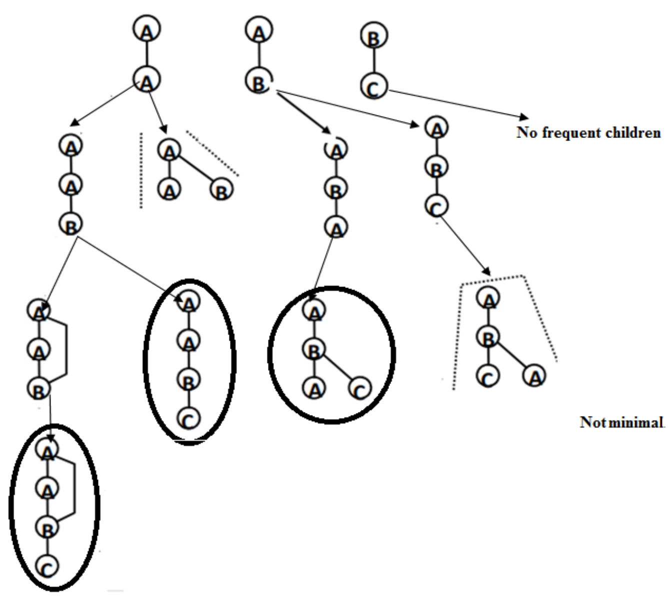


**Figure 8: Graph obtained using gSpan with minimal support =3**

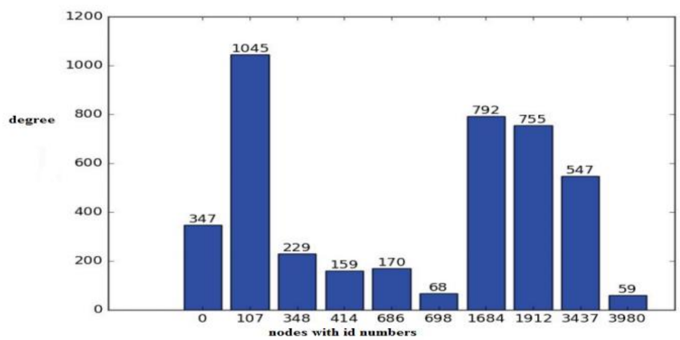
## 7. **Dataset**

In this paper in order to discover the trends in the social network like Facebook like social network we considered a data site which includes 10 ego-networks for different 10 facebook members.[7] From the dataset, ego nodes are defined by numbers called node id. Some of the ego node id's are 0, 107, 348, 414, 686, 698, 1684, 1912, 3437, 3980.

The number of nodes and edges of all ego networks are given in the table.

**Table 1: nodes and edges of all ego networks (small facebook data)**

The degree of each node gives the number of friendship connection.




We observed that a node (facebook user) with node id 107 is shows the largest ego network with 1045 connections and ego network 3980 is the smallest ego network with 59 connected nodes. We performed this test using gSpan algorithm. Community networks.

## 8.Results

| Algorithm | Input | Output | Graph |
|---|---|---|---|
| FSgM algorithm | Set of graphs | Frequent subgraph | Adjacency list |
| gSpna | Set of graphs | Frequent subgraph | Adjacency list |

Results obtained on a Small Ego Network from Facebook

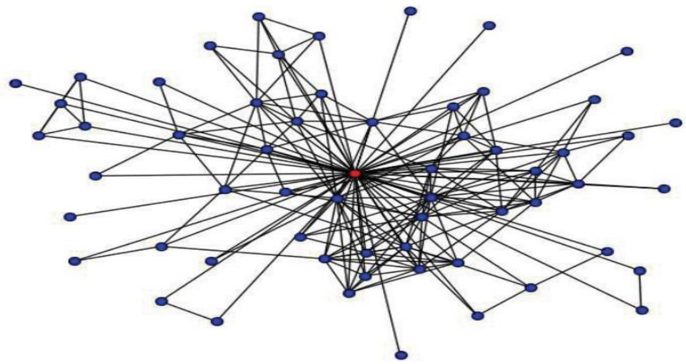


**Figure 9: Small Ego Network from Facebook**

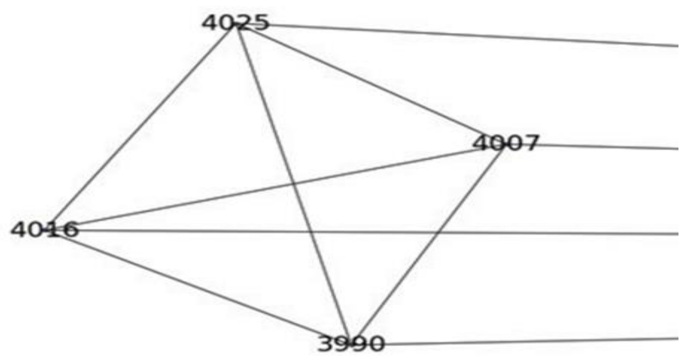



**Figure 10. Small part of Network 3980**

| Feature Set | Support | Patterns found |
|---|---|---|
| [ 14, 22, 53, 54, 55, 78, 127] | 7 | 807 |
| [ 22, 53, 54, 55, 78, 127] | 6 | 25756 |
| [14, 22, 53, 54, 55, 127] | 6 | 807 |
| [ 22, 53, 54, 55, 127] | 5 | 25756 |

**Table :** gSpan algorithm results on the smallest ego network 3980

## CONCLUSION

In today's internet world accessing to the Social networks has become a trend like a meeting point for today's internet savvy audience. Social media networks have become important role to people, for share activities and interests. This results to the spread of enormous amount of useful information, and opened a path to trending data mining, graph mining data with analysis offering lots of advantages to the research communities. The most efficient way to represent a social network data is by graphs. So Graph mining has become one of the most important disciplines in the research area of data mining. A Frequent Sub graph mining (FSgM) techniques has been proposed to identify the frequent pattern trends existing in the network social networks like facebook. We have successfully implemented the proposed algorithm gSpan for identifying interesting frequent pattern trends in facebook social network data. The final results have proved that our experimental works on small social ego network has most common frequent patterns in friendship relationships of Facebook dataset 78-78 and 127-127.

## REFERENCES

[1] S. Thomas and J. J. Nair, ‘‘A survey on extracting frequent subgraphs,’’ presented at the Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), 2016, pp. 2290–2295.

[2] S. Zhang, J. Yang, and V. Cheedella, ‘‘Monkey: Approximate graph mining based on spanning trees,’’ presented at the IEEE 23rd Int. Conf. Data Eng. (ICDE), 2007, pp. 1247–1249.

[3] Z. Zou, J. Li, H. Gao, and S. Zhang, ‘‘Mining frequent subgraph patterns from uncertain graph data,’’ IEEE Trans. Knowl. Data Eng., vol. 22, no. 9, pp. 1203–1218, Sep. 2010.

[4] Darren Rolfe + Vince Chu Graph Mining and Social Network analysis Data Mining; EECS 4412, 11.06.14

[5] T. Nguyen and P. Do, ‘‘Topic discovery using frequent subgraph mining approach,’’ in Proc. Int. Conf. Comput. Sci. Technol.. Singapore: Springer, Nov. 2017, pp. 432–442

[6] T. Alam, S. A. Zahin, M. Samiullah, and C. F. Ahmed, ‘‘An efficient approach for mining frequent subgraphs,’’ presented at the Int. Conf. Pattern Recognit. Mach. Intell., 2017, pp. 486–492.

[7] J. McAuley and J. Leskovec. Learning to Discover Social Circles in Ego Networks. NIPS, 2012.Snap

https://snap.stanford.edu/data/egonets-Facebook.html

[8] Frequent Subgraph Patterns (FSPs) mining (Borgelt and Berthold 2002; Kuramochi and Karypis 2001,

2004.

[9] Diane J. Cook and Lawrence B. Holder, Mining Graph Data, Electrical Engineering and Computer

Science, Washington State University, Pullman, Washington, Copyright _c 2007 John Wiley & Sons, Inc.
[10]Algur SP, Bhat P (2016) Web video object mining: expectation maxi- mization and density based clustering of web video metadata objects. Int J Inform Eng Electron Bus 8(1):69