

Available online at <http://www.mecspress.net/ijeme>

An Approach to Represent Social Graph as Multi-Layer Graph Using Graph Mining Techniques

Bapuji Rao *

Asst. Professor, Department of CSE&A, IGIT, Sarang, Dhenkanal-759146, India

Received: 10 July 2018; Accepted: 16 October 2018; Published: 08 January 2019

Abstract

In Social Graph, a set of entities or nodes or vertices interact with each other in a complicated manner that can form multiple types of relationships that depend on time and types of complications. Such graphs include multiple subsystems and layers of connectivity. So it is important to take such multi-layer features into account to make easier of understanding of such complex systems. In this paper, the author focuses on a Social Graph to represent in a multi-layer graph based on its characteristics lies in each node or vertex or entity. For this, the author proposes a general model related to Social Graph. For this model, the author proposes an algorithm, SoGraM for representation of Social Graph with multi-layer features using Graph Mining Techniques. Further, the author tries to prove the proposed algorithm with three examples of Social Graph namely Author Graph, Email Graph, and Telephone Graph.

Index Terms: Adjacency Matrix, Base-Layer, Multi-Layer Graph, Social Graph, Sub-Layer.

© 2019 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Multi-layer networks exist when there is more than one source of connectivity information for a group of users. In the context of the social network, the communication link between the users is relational information. So there is a possibility of deriving behavioral relationships based on user actions or interests [9].

In social networks, edges are categorized based on the nature of the relationships or actions that they represent are proposed by [6, 13]. In the social networks, the edge is categorized by its type are called multiplex networks by [4] or multi-relational networks proposed by [6]. In the computer science and computational linear-algebra communities, tensor-decomposition methods proposed by [3, 5] and multi-way data analysis by

* Corresponding author:
E-mail address:

[1] were used to study various types of multi-layer networks. These types of methods are based on representing multi-layer networks as adjacency tensors of order higher than two.

Networked systems that cannot be represented as traditional graphs were studied from a data mining point of view. Heterogeneous (information) networks were developed as a general framework to take into account multiple types of nodes and edges can be seen in [2, 12, 14]. A multiple layers of a social network which perform tasks such as inference, clustering, and anomaly detection proposed by [7]. Often social networks include different types of nodes i.e. males and females or hierarchical structures i.e. individuals are considered as organizations, were studied using multi-level networks, multiple layers, and networks of networks by [4]. The extraction of information from a large multi-layer social network is proposed by [8]. A multi-layer social network model which combines all the detected social network while encounters with other social networks are proposed by [11]. These multi-layer social networks are investigated through relationship.

A social network is defined as a complex graph, and each unit is an individual, village, household, country, etc. So the social network may be defined as set nodes or vertices $V = \{V_1, V_2, \dots, V_n\}$ and between the nodes or vertices there is a set of links or connectivity or edges $E = \{E_1, E_2, \dots, E_n\}$ proposed by [9]. In this paper, the author proposes an algorithm which represents a social graph as a multi-layer graph using graph mining techniques. For this, the authors represent three types of social graph examples namely Authors Graph, Email Graph, and Telephone Graph that consist of multi-layer graphs where the base-layer or principal layer include different type of interactions between entities or nodes or vertices. An algorithm which represents telephone call duration graph as a multi-layer graph is proposed by [10]. Based on [10]'s algorithm, the author proposes a general algorithm for representation of the social graph as a multi-layer graph having each level is associated with a particular characteristic.

2. Multi-layer Network

In the real world, more than one kind of connections can exist between any pair of individuals. As a matter of fact, for this type of social networks, it can be termed as the multi-layer network, multi-relational network, multidimensional network, and multiplex network.

A social network layer L is a weighted graph $G(V, E)$ with vertex set V corresponding to users on the social network and edge set $E \subseteq VXV$ corresponding to social links between users.

A multi-layer social network $MLSN = (L_1, L_2, \dots, L_n)$ is a tuple where $L_i = G_i(V, E_i)$, $i \in 1, 2, \dots, n$ are social network layers.

A multi-layer network $G = (V, E)$ consists of vertices $V = \{V_1, \dots, V_n\}$, common to all layers, and edges $E = (E_1, \dots, E_m)$ in m layers, where E_k is the edge set for layer k , and $E_k = \{e_{ij}^k; V_i, V_j \in V\}$. Each edge is considered as directed or undirected depending on its characteristics. The definition of multi-layer network can be seen in [8].

3. Proposed Algorithm for Representation of Social Graph as Multi-layer Graph

The proposed algorithm, SoGraM has only one procedure called Show(MLAM). The algorithm reads the data from the dataset file called "Node_Dataset.Txt" which comprises a number of layers, total number of nodes, and all node Ids. The first and second data from the dataset file are considered as the number of layers and the total number of nodes, which is assigned to the variables NL and n respectively. Then from third data onwards are considered as the node IDs. It starts reading one ID at a time and assigned to the variable $Node_Id$. Finally, the variable $Node_Id$ is assigned to the row header and column header of the $MLAM[][]$'s principal layer or base-layer. After the complete reading of all node IDs, the initial form of the multi-dimension array, $MLAM[][]$ is formed successfully.

Then the second dataset file called "Edge_Dataset.Txt" which comprises all edges ID in the forms of two column headers 'From Node Ids' and 'To Node Ids'. Then it starts reading at a time a pair of node IDs and assigned to the variables $From_Node$ and To_Node respectively. So the particular social graph's edge data is

assigned successfully on the first layer of the multi-dimension array, $MLAM[][][]$. Hence it is the indication of the formation of the principal layer or base-layer of the social graph. By using the multi-dimension array, $MLAM[][][]$ which has a principal layer at the topmost layer, the required number of layers i.e. (NL-1) numbers are formed starting from 2nd layer to (NL-1) layer based on the criteria. So finally, the multi-dimension array, $MLAM[][][]$ with all layers assigned with the proper edge values i.e. either 1s or 0s.

Finally, the procedure Show(MLAM) passes the multi-dimension array, $MLAM[][][]$ and displays the multi-dimension array, Matrix with a number of layers, b numbers of rows, and c numbers of columns respectively.

i. Algorithm SoGram()

n: Number of nodes in the social graph.

NL: Number of layers.

$MLAM[NL][n+1][n+1]$: Multi-dimension adjacency matrix to store the social graphs multi-layers edge data.

Node_Dataset.Txt: Text files to contain number of layers, total number of nodes, and node Ids.

Edge_Dataset.Txt: Text files to contain 'From Node Ids' and 'To Node Ids' where the actual edge is present.

```
{
// creation of adjacency matrix of social graph
open("Node_Dataset.Txt");
read(NL, n); // to read number of layers and number of nodes
for i:=2 to (n+1) do
{
  Read(Node_Id); // to read node ID
   $MLAM[1][i]:=MLAM[i][1]:=Node\_Id$ ;
}
close("Node_Dataset.Txt");
open("Edge_Dataset.Txt");
while(Not EOF())
{
  read(From_Node, To_Node); // to read 'from node id' and 'to node id'
  for i:=2 to (n+1) do {
    if( $MLAM[1][i][1]=From\_Node$ ) then break; // to detect the row node id }
  for j:=2 to (n+1) do {
    if( $MLAM[1][1][j]=To\_Node$ ) then break; // to detect the column node id }
   $MLAM[1][i][j]:=1$ ; // to assign edge value 1 at ith row and jth column in  $MLAM[1][ ][ ]$ 
}
close("Edge_Dataset.Txt");
// creation of 2nd layer to NL layers
for i:=2 to NL do
{
  for j:=1 to n do
  for k:=1 to n do
  {
    if(Criteria of ith layer on  $MLAM[1][j][k]$  is TRUE) then
       $MLAM[i][j][k]:=MLAM[1][j][k]$ ;
  }
}
}
Show(MLAM);
}
```

ii. Procedure to Display Multi-layer Graphs Adjacency Matrix

Procedure Show(Matrix[a][b][c])

a: Number of Layers.

b: Number of Rows.

c: Number of Columns.

```
{
  for i:=1 to a do
  {
    display("Layer-", i);
    for j:=1 to b do {
      for k:=1 to c do {
        display(Matrix[i][j][k]); } }
  }
}
```

4. Examples of Proposed Algorithm

The author has considered three social graphs namely Author Graph, Email Graph, and Telephone Graph to represent as multi-layer graphs. For Author Graph, the total numbers of author nodes are 22 having with node id ranging from 1 to 22. It is to represent with three numbers of layers having considered the 1st layer as the base-layer (Layer-1) followed by two layers namely Layer-2 and Layer-3 respectively. The second social graph example is the Email Graph. This graph has total numbers of email ids are 22 having with node id ranging from 1 to 22. The author is to represent the email graph with four numbers of layers having considered the 1st layer as the base-layer (Layer-1) followed by three layers namely Layer-2, Layer-3, and Layer-4 respectively. Finally, the author is to represent Telephone Graph as a multi-layer graph having with 22 telephone node ids ranging from 11 to 17, 22 to 27, and 33 to 38 respectively. In node id, 1 represents service provider 1, 2 represents service provider 2, and 3 represents service provider 3 respectively. This graph is to represent with five numbers of layers having considered Layer-1 is the base-layer and followed by four numbers of layers namely Layer-2, Layer-3, Layer-4, and Layer-5 respectively.

i. Representation of Author Graph as Multi-layer Graph

The author has considered the first example as Author Graph is a social graph example which has twenty-two number of author nodes ranging from 1 to 22. This authors graph is a directed graph since there is a kind of participation of authors in the publication of a paper and depicted in "Fig. 1". The actual author graph is represented on the 1st layer is the principal or base-layer. Then the required two numbers of subsequent layers are to be generated from base-layer based on some criteria. The 2nd layer is generated where the criteria is a publication of paper jointly. So a sub-graph is generated on the 2nd layer which is considered as a joint-authors publication sub-graph. To detect those authors joint participation in the publication of a paper is the bi-directional edge between the nodes in the base or 1st layer's graph and such detected nodes edges are shown in the 2nd layer as a sub-graph. Similarly, the 3rd layer is generated where the criteria are the publication of paper as the first author. So a sub-graph is generated on the 3rd layer is the publication of paper as the first author. To detect those nodes in the 1st layer graph, the node has no out-degree but only an in-degree and such detected nodes edges are shown in the 3rd layer as a sub-graph. So in the 3rd layer's sub-graph, the black filled nodes 6, 8, 18, and 22 are considered as the first author of the publication of a paper.

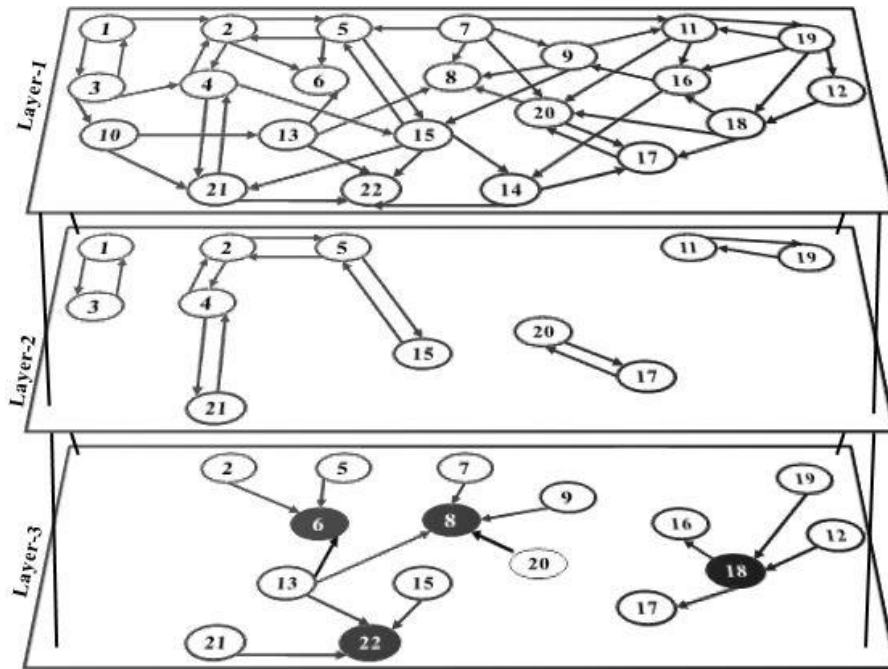


Fig.1. Author Multi-layer Graph

ii. Representation of Email Graph as Multi-layer Graph

The author has considered the second example as Email Graph is one of the social graph examples. It has twenty-two numbers of email nodes ranging from 1 to 22. This email graph is also a directed graph since there is a kind of participation of email users in the sending of emails for a period of time and depicted in "Fig. 2". The actual email graph is represented on the 1st layer is considered as the principal or base-layer. Then the required three numbers of subsequent layers are to be generated from base-layer based on some criteria which are suited for the email graph. The 2nd layer is generated where the criteria are the mutual communication between the email users. So a sub-graph is generated on the 2nd layer which is considered as mutual communicator sub-graph. To detect those email nodes from the base or 1st layer email graph, there must be a bi-directional edge between the email nodes and such email nodes edges are shown in the 2nd layer as a sub-graph. Similarly, the 3rd layer is generated where the criteria is a one-way communication between the email users. So a sub-graph is generated on the 3rd layer which is considered as a one-way communicator between the emails nodes. To detect those email nodes from the base or 1st layer email graph, there must be the only unidirectional edge which is pointing away to other email nodes i.e. the email node is having only out-degree and such nodes edges are shown in the 3rd layer as a sub-graph. Then the 4th and final layer is generated where the criteria are the no communication is sent from the email nodes among other email nodes in the email graph. So a sub-graph is generated on the 4th layer which is considered as dead-communicator sub-graph. To detect those email nodes from the base or 1st layer email graph, there must be only edges pointing to the email node from other email nodes i.e. the email node has no out-degree and such nodes edges are shown in the 4th layer as sub-graph. However, the nodes with only in-degree but no out-degree are considered as dead communicators for a period of time. In "Fig. 2", the black filled nodes 8, 17, and 22 are considered as the dead communicators.

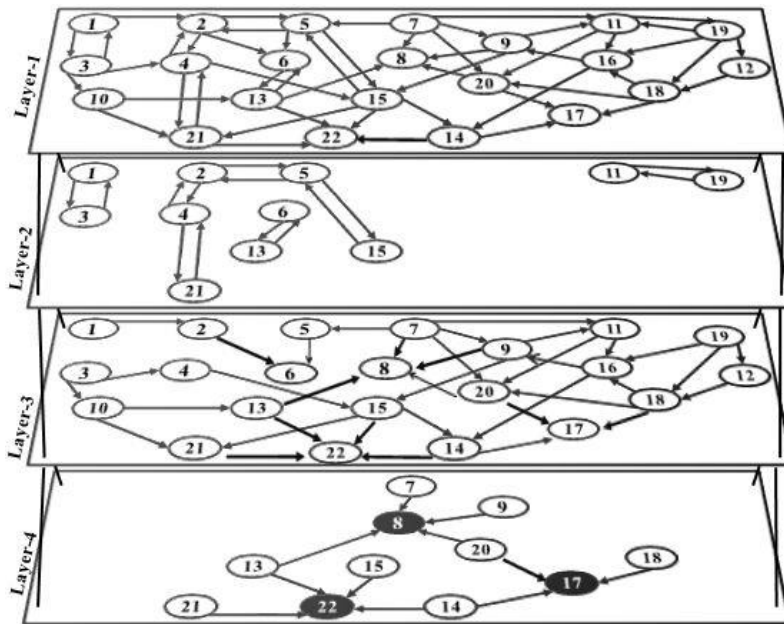


Fig.2. Email Multi-layer Graph

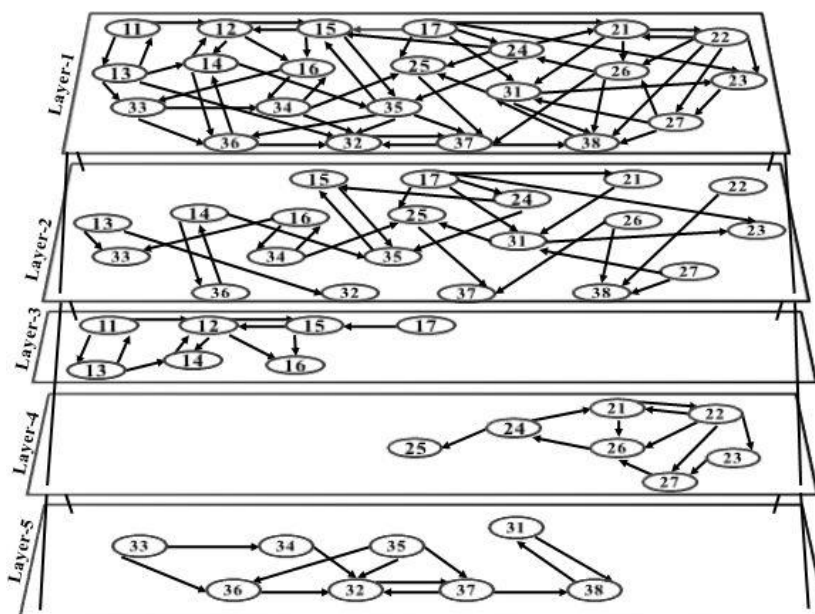


Fig.3. Telephone Multi-layer Graph

iii. Representation of Telephone Graph as Multi-layer Graph

The third example of the social graph is the Telephone Graph. It consists of twenty-two numbers of telephone number nodes ranging from 11 to 17, 22 to 27, and 33 to 38 respectively. It is a directed graph and

the indication of directed edge is the participation of a telephone number node calling to other telephone number node in the telephone graph and depicted in “Fig. 3”. The 1st digit of node ID refers to the service provider codes. Therefore, there are three service providers having service provider codes 1, 2, and 3 respectively. The actual telephone graph is represented on the 1st layer is considered as the principal or base-layer. The author needs to generate four numbers of layers based on some criteria. The 2nd layer is generated where the criteria are of dissimilar service provider’s calling made by the telephone user in the telephone graph. So a sub-graph is generated on the 2nd layer where the calls are made between dissimilar service providers and the sub-graph is considered as dissimilar service provider’s sub-graph. To detect those nodes from the base or principal layer telephone graph, there must be an edge between two nodes from different service providers and such telephone nodes edges are shown in the 2nd layer as a sub-graph. Then the 3rd, 4th, and 5th layers for generation of the same service provider’s sub-graphs since the telephone graph have three types of service providers. Based on the criteria these service providers nodes are detected from the base or principal layer telephone graph and corresponding nodes are shown in the layers 3rd, 4th, and 5th respectively as sub-graphs. Therefore, these sub-graphs are considered as 1st service provider sub-graph, 2nd service provider graph, and 3rd service provider graph respectively.

5. Experimental results

The author has considered three examples of social graphs namely Author Graph, Email Graph, and Telephone Graph. To implement the algorithm for the above three graphs, the author has created three sets of datasets namely {Anode.Txt, Aedge.Txt} for Author Graph, {Enode.Txt, Eedge.Txt} for Email Graph, and {Tnode.Txt, Tedge.Txt} for Telephone Graph respectively.

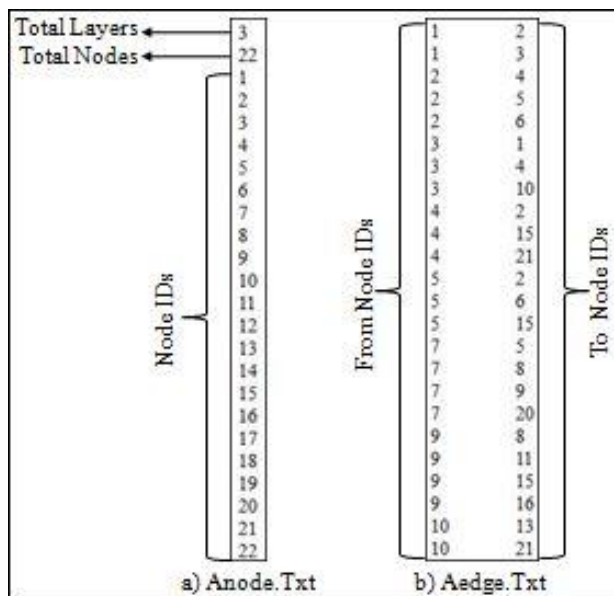


Fig.4. Dataset of Author Graph

i. Example-1

The Author Multi-Layer graph has twenty-two numbers of author nodes and three layers of graphs based on their characteristics depicted in “Fig. 1”. To run the proposed algorithm, the author has created two datasets

namely “Anode.Txt” and “Aedge.Txt” depicted in “Fig. 4”. The 1st dataset “Anode.Txt” consists of the total number of layers, the total number of nodes, and the twenty-two numbers of node IDs. The 2nd dataset “Aedge.Txt” contains a pair of nodes “From Node IDs” and “To Node IDs” where the actual edge is present in the graph. The dataset files “Anode.Txt” and “Aedge.Txt” are replaced with “Node_Dataset.Txt” and “Edge_Dataset.Txt” in the algorithm for representation of Authors Graph as multi-layer graph with three layers based on its characteristics. To represent Authors Graph as a multi-layer graph in memory, the author has considered a multi-dimension array MLAM[3][22][22], with three numbers of layers. Here each layer is considered as adjacency matrix having order 22X22.

Upon inputting the datasets “Anode.Txt” and “Aedge.Txt” to the algorithm depicted in “Fig. 5”, it has successfully created three layers.

The 1st layer depicted in “Fig. 6” is considered as base-layer or principal layer. The subsequent two layers are formed depending on its criteria.

The 2nd layer is formed based on the criteria of detection of author nodes having the joint-authors publication of paper i.e. between a pair of author node there must be a bi-directional edge. Such author node pairs are detected from the 1st layer adjacency matrix and assign a value 1 in the 2nd layer’s corresponding nodes of the adjacency matrix. Here the author node pairs are {(1, 3), (3, 1)}, {(2, 4), (4, 2)}, {(2, 5), (5, 2)}, {(4, 21), (21, 4)}, {(5, 15), (15, 5)}, {(11, 19), (19, 11)}, and {(17, 20), (20, 17)} which has bi-directions with assignments of 1s and depicted in “Fig. 1 (Layer-2)” and “Fig. 7” respectively.

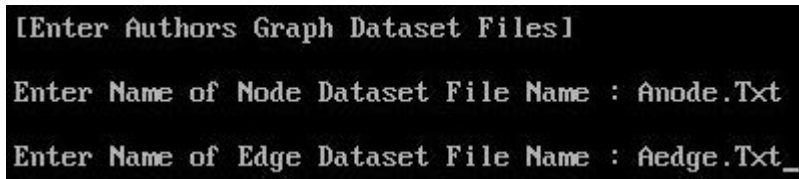


Fig.5. Dataset input of Authors Graph

Layer-1 of Author Graph																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
5	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0
20	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Press Any Key To See The Next Layer..._

Fig.6. Adjacency Matrix of Authors Graph

Layer-2 of Author Graph

2	1	2	3	4	5	11	15	17	19	20	21
1	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	1	1	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	1
5	0	1	0	0	0	0	1	0	0	0	0
11	0	0	0	0	0	0	0	0	1	0	0
15	0	0	0	0	1	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	1	0
19	0	0	0	0	0	1	0	0	0	0	0
20	0	0	0	0	0	0	0	1	0	0	0
21	0	0	0	1	0	0	0	0	0	0	0

Press Any Key To See The Next Layer..._

Fig.7. Adjacency Matrix of Multiple Authors' Publication of Papers Sub-graph

Layer-3 of Author Graph

3	6	8	14	18	22
2	1	0	0	0	0
5	1	0	0	0	0
7	0	1	0	0	0
9	0	1	0	0	0
12	0	0	0	1	0
13	1	1	0	0	1
15	0	0	1	0	1
16	0	0	1	1	0
17	0	0	1	1	0
19	0	0	0	1	0
20	0	1	0	0	0
21	0	0	0	0	1

Fig.8. Adjacency Matrix of First Author Publication of Papers Sub-graph

Similarly, the 3rd layer is formed based on the criteria of detection of author node which is responsible as the first author for publication of papers. The author nodes having only in-degree but no out-degree are detected from the 1st layer adjacency matrix and assigned a value 1 to all the corresponding nodes in the 3rd layer adjacency matrix which is the indication of in-degrees. Here the author nodes with out-degree zero are 6, 8, 14, 18, and 22 depicted in "Fig. 1 (Layer-3)" and "Fig. 8" respectively.

ii. Example-2

The Email Multi-Layer graph has twenty-two numbers of email nodes and four layers of graphs based on their characteristics depicted in “Fig. 2”. To run the proposed algorithm, the author has created two datasets namely “Enode.Txt” and “Eedge.Txt” depicted in “Fig. 9”. The 1st dataset “Enode.Txt” consists of the total number of layers, the total number of nodes, and the twenty-two numbers of node IDs. The 2nd dataset “Eedge.Txt” contains a pair of email nodes “From Node IDs” and “To Node IDs” where the actual edge is present in the graph.

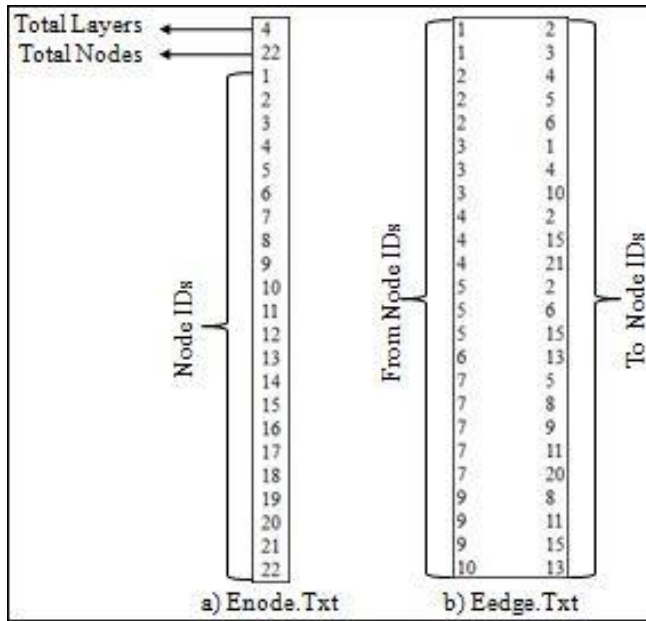


Fig.9. Dataset of Email Graph

The dataset files “Enode.Txt” and “Eedge.Txt” are replaced with the “Node_Dataset.Txt” and “Edge_Dataset.Txt” in the algorithm for representation of Email Graph as a multi-layer graph with four layers based on its criteria.

To represent the Email Graph as a multi-layer graph in memory, the author has considered a multi-dimension array MLAM[4][22][22], with four numbers of layers and each layer, is considered as the adjacency matrix of order 22X22.

Upon inputting the datasets “Enode.Txt” and “Eedge.Txt” to the algorithm depicted in “Fig. 10”, it has successfully created four layers and depicted in “Fig. 11”, “Fig. 12”, “Fig. 13”, and “Fig. 14” respectively.

```
[Enter Email Graph Dataset Files]
Enter Name of Node Dataset File Name : Enode.Txt
Enter Name of Edge Dataset File Name : Eedge.Txt_
```

Fig.10. Dataset Input of Email Graph

Layer-1 of Email Graph

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
5	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
15	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
16	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
19	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0
20	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Press Any Key To See The Next Layer..._

Fig.11. Adjacency Matrix of Email Graph

Layer-2 of Email Graph

	2	3	4	5	6	11	13	15	19	21
2	1	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0
5	0	1	0	1	0	0	0	1	0	0
6	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	1	0	0	0	0	0
15	0	0	0	1	0	0	0	0	0	0
19	0	0	0	0	0	1	0	0	0	0
21	0	0	1	0	0	0	0	0	0	0

Press Any Key To See The Next Layer..._

Fig.12. Adjacency Matrix of Two-way Communicator Sub-graph

Layer-3 of Email Graph																			
	3	2	4	5	6	8	9	10	11	12	13	14	15	16	17	18	20	21	22
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
16	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
19	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
20	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Press Any Key To See The Next Layer...

Fig.13. Adjacency Matrix of One-way Sender Sub-graph

Layer-4 of Email Graph				
	4	8	17	22
7	1	0	0	0
9	1	0	0	0
13	1	0	1	0
14	0	1	1	0
15	0	0	1	0
18	0	1	0	0
20	1	1	0	0
21	0	0	1	0

Fig.14. Adjacency Matrix of One-way Receiver Sub-graph

The 1st layer is considered as base-layer or principal layer. The subsequent three layers are formed depending on its criteria.

The 2nd layer depicted in “Fig. 12” is formed based on the criteria of detection of mutual (two-way) communicator email node from the 1st layer adjacency matrix and the corresponding node pairs are assigned with a value 1 in the 2nd layer adjacency matrix. Here the email node pairs {(1, 3), (3, 1)}, {(2, 4), (4, 2)}, {(2, 5), (5, 2)}, {(4, 21), (21, 4)}, {(5, 15), (15, 5)}, {(6, 13), (13, 6)}, and {(11, 19), (19, 11)} are considered as two-way communicators and depicted in “Fig. 2 (Layer-2)” and “Fig. 12” respectively.

Similarly, the 3rd layer is formed based on the characteristics of detection of email nodes which has sent emails to other email nodes but not received any emails during a period of time. Such email nodes edges are detected from the 1st layer adjacency matrix and the corresponding nodes are assigned a value 1 in the 3rd layer adjacency matrix. It means detection of only such mail nodes having edge away from it and depicted in “Fig. 2 (Layer-3)” and “Fig. 13” respectively. The node pairs (1, 2), (2, 6), (3, 4), (3, 10), (4, 15), (5, 6), (7, 5), (7, 8),

(7, 9), (7, 11), (7, 20), (9, 8), (9, 11), (9, 15), (10, 13), (10, 21), (11, 16), (11, 20), 12, 18), (13, 8), (13, 22), (14, 17), (14, 22), (15, 14), (15, 21), (15, 22), (16, 9), (16, 14), (18, 16), (18, 17), (18, 20), (19, 12), (19, 16), (19, 18), (20, 8), (20, 17), and (21, 22) are considered as one-way sender and the adjacency matrix formed is said to be one-way sender adjacency matrix.

Finally, the 4th layer is formed based on the characteristics of detection of email node which has received emails but not sent emails during a period of time. So the email nodes having only in-degree but no out-degree are detected from the 1st layer adjacency matrix and the corresponding nodes are assigned a value 1 in the 4th layer adjacency matrix. The created adjacency matrix is said to be one-way receiver adjacency matrix. Here the email nodes 8, 17, and 22 are said to be dead email nodes since it has received emails from other email nodes and depicted in “Fig. 2 (Layer-4)” and “Fig. 14” respectively.

iii. Example-3

The Telephone Multi-Layer graph has twenty-two numbers of phone nodes and five layers of graphs based on their characteristics depicted in “Fig. 3”. To run the proposed algorithm, the author has created two datasets namely “Tnode.Txt” and “Tedge.Txt” depicted in “Fig. 15”. The 1st dataset “Tnode.Txt” contains the total number of layers, the total number of nodes, and the twenty-two numbers of phone node IDs. The 2nd dataset “Tedge.Txt” contains a pair of phone nodes “From Node IDs” and “To Node IDs” where the actual edge is present in the graph.

The dataset files “Tnode.Txt” and “Tedge.Txt” are replaced with “Node_Dataset.Txt” and “Edge_Dataset.Txt” in the algorithm for representation of Telephone Graph as multi-layer graph with five layers based on its criteria.

To represent Telephone Graph as a multi-layer graph in memory, the author has considered a multi-dimension array MLAM[5][22][22], with five numbers of layers and each layer, is considered as the adjacency matrix of order 22X22.

Upon inputting the datasets “Tnode.Txt” and “Tedge.Txt” to the algorithm depicted in “Fig. 16”, which has successfully created five layers of adjacency matrices.

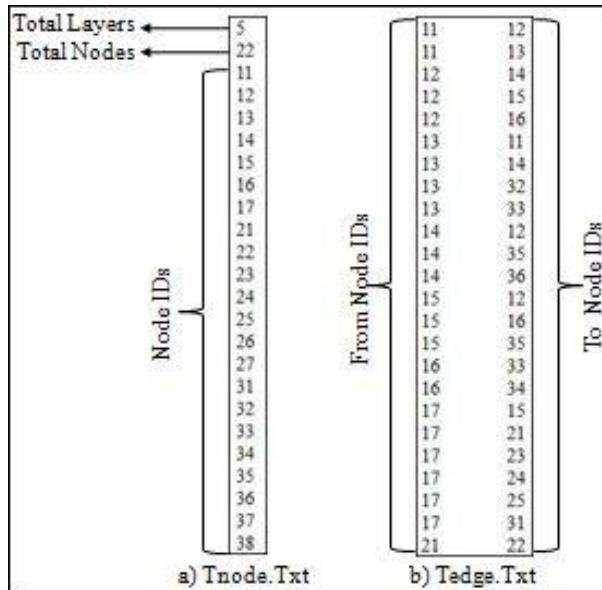


Fig.15. Dataset of Telephone Graph

```
[Enter Telephone Graph Dataset Files]
Enter Name of Node Dataset File Name : Tnode.Txt
Enter Name of Edge Dataset File Name : Tedge.Txt
```

Fig.16. Dataset Input of Telephone Graph

```
Layer-1 of Telephone Graph
 1 11 12 13 14 15 16 17 21 22 23 24 25 26 27 31 32 33 34 35 36 37 38
11 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
12 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
13 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
14 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0
15 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
16 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
17 0 0 0 0 1 0 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 0
21 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0
22 0 0 0 0 0 0 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 1
23 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
24 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0
25 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
26 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1
27 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1
31 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1
32 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
33 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
34 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0
35 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
36 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
37 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1
38 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
Press Any Key To See The Next Layer...
```

Fig.17. Adjacency Matrix of Telephone Graph

```
Layer-2 of Telephone Graph
 2 14 15 16 21 23 24 25 31 32 33 34 35 36 37 38
13 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0
14 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0
15 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
16 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0
17 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0
21 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
22 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
24 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0
25 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
26 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
27 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1
31 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0
34 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0
35 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
36 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Press Any Key To See The Next Layer..._
```

Fig.18. Adjacency Matrix of Dissimilar Service Provider's Connectivity Sub-graph

The 1st layer depicted in “Fig. 17” is considered as the base-layer or principal layer. The remaining four layers are formed depending on its criteria.

The 2nd layer depicted in “Fig. 18” is considered as dissimilar service providers’ communication sub-graph. It has detected the phone node pair connectivity between two different service providers. Such phone node pairs are detected from base-layer adjacency matrix and the corresponding node pairs are assigned a value 1 in the 2nd layer adjacency matrix. The node pairs (13, 32), (13, 33), (14, 35), (14, 36), (15, 35), (16, 33), (16, 34), (17, 21), (17, 23), (17, 24), (17, 25), (17, 31), (21, 31), (22, 38), (24, 15), (24, 35), (25, 37), (26, 37), (26, 38), (27, 31), (27, 38), (31, 23), (31, 25), (34, 16), (34, 25), (35, 15), and (36, 14) are considered as different service providers communication link.

Similarly, the 3rd, 4th, and 5th layers are formed from the 1st layer by following the criteria as service provider 1, service provider 2, and the service provider 3 connectivity and depicted in “Fig. 19”, “Fig. 20”, and “Fig. 21” respectively. The 3rd layer is created from 1st layer adjacency matrix by detecting the nodes of service provider 1’s connectivity and the corresponding nodes in the 3rd layer adjacency matrix are assigned with a value 1. The service provider 1’s node ids ranging from 11 to 17 and the pair of nodes whose connectivity among themselves are {(11, 12), (11, 13), (12, 14), (12, 15), (12, 16), (13, 11), (13, 14), (14, 12), (15, 12), (15, 16), (17, 15)}. Similarly, the 4th and the 5th layers are created from 1st layer adjacency matrix by detecting the nodes of the service provider 2’s and service provider 3’s connectivity, and the corresponding nodes in 4th and 5th layer adjacency matrices are assigned a value 1. The service provider 2’s node ids ranging from 21 to 27 and the pair of nodes whose connectivity among themselves are {(21, 22), (21, 26), (22, 21), (22, 23), (22, 26), (22, 27), (23, 27), (24, 21), (24, 25), (26, 24), (27, 26)}. Similarly, the service provider 3’s node ids ranging from 31 to 38 and the pair of connectivity nodes are {(31, 38), (32, 37), (33, 34), (33, 36), (34, 32), (35, 36), (35, 37), (36, 32), (37, 32), (37, 38), (38, 31)}.

```

Layer-3 of Telephone Graph
 3 11 12 13 14 15 16
11 0 1 1 0 0 0
12 0 0 0 1 1 1
13 1 0 0 1 0 0
14 0 1 0 0 0 0
15 0 1 0 0 0 1
17 0 0 0 0 1 0
Press Any Key To See The Next Layer...
    
```

Fig.19. Adjacency Matrix of Service Provider 1 Sub-graph

```

Layer-4 of Telephone Graph
 4 21 22 23 24 25 26 27
21 0 1 0 0 0 1 0
22 1 0 1 0 0 1 1
23 0 0 0 0 0 0 1
24 1 0 0 0 1 0 0
26 0 0 0 1 0 0 0
27 0 0 0 0 0 1 0
Press Any Key To See The Next Layer..._
    
```

Fig.20. Adjacency Matrix of Service Provider 2 Sub-graph

Layer-5 of Telephone Graph						
5	31	32	34	36	37	38
31	0	0	0	0	0	1
32	0	0	0	0	1	0
33	0	0	1	1	0	0
34	0	1	0	0	0	0
35	0	0	0	1	1	0
36	0	1	0	0	0	0
37	0	1	0	0	0	1
38	1	0	0	0	0	0

Fig.21. Adjacency Matrix of Service Provider 3 Sub-graph

6. Conclusion

The author has proposed a general model of Social Graph for representing in a multi-layer graph based on its characteristics lies in each node. For this general model of social graph multi-layer representation, the author has proposed an algorithm called SoGraM for representation of Social Graph as a multi-layer graph. Finally, the proposed algorithm was implemented using C++ programming with three examples of datasets depicted in “Fig. 4”, “Fig. 9”, and “Fig. 15” respectively. The author has observed the formations of three layers, four layers, and five layers for Author Graph, Email Graph, and Telephone Graph respectively as multi-dimensional arrays in the memory and the results were satisfactory.

References

- [1] Acar, E. & Yener, B. Unsupervised multiway data analysis: a literature survey. *IEEE Trans. Knowl. Data Eng.*, 21, 6–20, 2009.
- [2] Cai, D., Shao, Z., He, X., Yan, X., & Han, J. Community mining from multi-relational networks. *Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Porto, Portugal, 2005.
- [3] Dunlavy, D. M., Kolda, T. G., & Kegelmeyer, W. P. Multilinear algebra for analyzing data with multiple linkages. *Graph Algorithms in the Language of Linear Algebra* (J. Kepner & J. Gilbert eds). *Fundamentals of Algorithms*. Philadelphia: SIAM, pp. 85–114, 2011.
- [4] Kivelä M., Arenas, A., Barthelemy, M., Gleeson, James P., Moreno, Y., & Porter, Mason A. Multilayer networks. *J. Complex Netw.*, 2, 203–271, 2014.
- [5] Kolda, T. G., & Bader, B. W. Tensor decompositions and applications. *SIAM Rev.*, 51, 455–500, 2009.
- [6] Krackhardt, D. Cognitive social structures. *Soc. Netw.*, 9, 109–134, 1987.
- [7] Oselio, B., Kulesza, A., & Hero, A.O. Multi-layer graph analysis for social networks. University of Michigan, Ann Arbor, MI 48109, USA. Available as arXiv:1309.5124v2, 2014.
- [8] Oselio, B., Kulesza, A., & Alfred Hero., A. O. Information extraction from large multi-layer social networks. *Proceedings of IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane. Available as arxiv 1507.00087, 2015.
- [9] Rao, B., Mitra, A., & Narayana, U. An approach to study properties and behavior of Social Network using Graph Mining Techniques. In the *Proceedings of DIGNATE 2014: ETEECT 2014*, 1–6, India, 2014.
- [10] Rao, B., Mishra, S. N., & Maharana, H. S. Algorithm for Representation of Call-Duration Graphs of Telephone Graph Using Multi-layer Graph Techniques. Published in *Advances in Intelligent Systems and Computing*, pp. 415-425, Springer Verlag, DOI 10.1007/978-981-10-2525-9, 2017.

- [11] Socievole, A., Yoneki, E., Rango, F. De., & Crowcroft, J. ML-SOR: Message routing using multi-layer social networks in opportunistic communications. Elsevier Computer Networks. <http://dx.doi.org/10.1016/j.comnet.2015.02.016>, 2015.
- [12] Sun, Y. & Han, J. Mining heterogeneous information networks: a structural analysis approach. ACM SIGKDD Explor. Newslett., 14, 20–28, 2013.
- [13] Wasserman, S. & Faust, K. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.
- [14] Zhou, D., Orshanskiy, S. A., Zha, H. & Giles, C. L. Co-ranking authors and documents in a heterogeneous network. Proceedings of Seventh IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, pp. 739–744, 2007.

Authors' Profile



Bapuji Rao is currently working as Asst. Professor in the Department of Computer Science, Engg. & Applications at Indira Gandhi Institute of Technology (IGIT), Sarang, Dhenkanal, Odisha, India. He is pursuing Ph.D. (CSE) from Biju Patnaik University of Technology (BPUT), Rourkela, Odisha, India. He has received M.Tech (Computer Science) from Berhampur University, Berhampur, Odisha, India. He has published 17 papers in International Conferences which includes IEEE-6, Springer-4, Elsevier-1, and McGraw-Hill-1. He has published 8 papers in international journals of repute. He has also published four book chapters in IGI Global, USA. His current research area focuses on Graph Mining, Data Mining, Opinion Mining, Text Mining, Social Network, Attributed Graph, and Multi-Layer Graph.

How to cite this paper: Bapuji Rao, "An Approach to Represent Social Graph as Multi-Layer Graph Using Graph Mining Techniques", International Journal of Education and Management Engineering(IJEME), Vol.9, No.1, pp.20-36, 2019.DOI: 10.5815/ijeme.2019.01.03