IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

Application of Graph Theory in Social Media

Anwesha Chakraborty[1*], Trina Dutta[2], Sushmita Mondal[3], Asoke Nath[4]

[1,2,3,4]Department of Computer Science, St. Xavier's College (Autonomous), Kolkata, India

**Corresponding Author: nancyc0087@gmail.com*

Available online at www.ijcseonline.org

Accepted: 22/Oct/2018, Published: 31/Oct/2018

Abstract**—** A graph is made up of nodes; just like that a social media is a kind of a social network, where each person or organization represents a node. These nodes in a social media are interdependent on each other via common interests, relations, mutual friends, knowledge, common dislikes, beliefs etc. The overall graphical structure of a social media can be very complex with millions of nodes and thousands of interconnections amongst them based upon various grounds. Many researchers have revealed that social network works on various levels and helps in understanding many things such as how an entire organization is run. It helps to solve and understand many critical problems.

The analysis of the social media is a very useful tool for extracting knowledge from unstructured data. The knowledge obtained from this field provides a vivid knowledge of various kinds interactions and relations amongst various individuals on social media. The authors have elaborated on the various applications of graph theory on social media and how it is represented viz. strong and weak ties. [1]

Keywords— Graph, Nodes, Social Media, Graphical structure, Unstructures data, Strong and Weak ties.

I. INTRODUCTION

What is a graph?

A graph represents a network which consists of a set of objects, mathematically called vertices or nodes. These vertices or nodes are interconnected with each other based upon some relation, with the help of nodes or arcs. [2]

Graph can be of two types based upon the type of edges:

i. **Directed Edges:** Here the arcs between two vertices have a particular direction; they are directed from one vertex to another. It is usually represented by an arrow.
ii. **Undirected Edges:** Here the edges do not have any particular direction from one vertex to another; there is no difference between the two vertices connected via one undirected edge. It is usually represented by a straight line.

A graph can again be classified based upon the presence or absence of cycles formed in the graph:

i. **Cyclic Graph:** If a cycle is formed in a graph then it is called a cyclic graph. It is a closed figure which starts and ends with the same vertex.
ii. **Acyclic Graph:** If the graph forms no cycle then it is called an acyclic graph. [3]

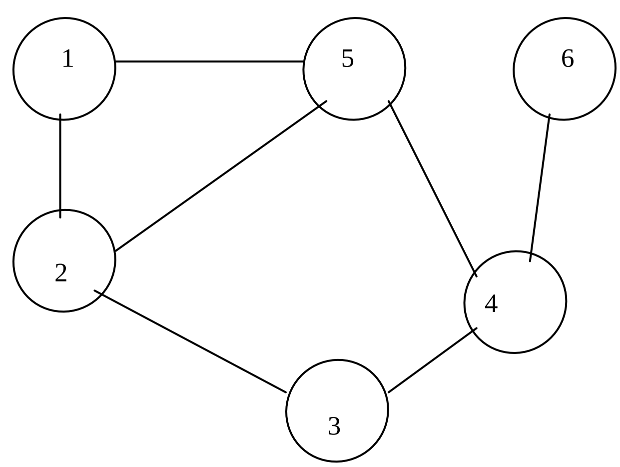
Representation of graphs

A graph can be represented mainly as two ways:

i. Adjacency matrices.
ii. Adjacency lists.

i. **Adjacency Matrices:** Here the graph is represented as an n X n square matrix; M. n represents the number of vertices present in the graph.

If $M_{ij} = 1$, it means there is an edge connecting vertex i and vertex j and if $M_{ij} = 0$, it means there is no edge connecting vertex i and vertex j. Let us consider the following 6x6 matrix:



	1	**2**	**3**	**4**	**5**	**6**
1	0	1	0	0	1	0
2	1	0	1	0	1	0
3	0	1	0	1	0	0
4	0	0	1	0	1	1
5	1	1	0	1	0	0
6	0	0	0	1	0	0

Figure 1: Adjacency Matrix and its corresponding graph.

Although the computation process in adjacency matrix is quite simple but it contains lots of zeroes and wastes a lots of space. In adjacency list representation of graphs, this disadvantage has been eliminated.

Adjacency list: In this case, all the zeroes of the adjacency matrix are eliminated and only the corresponding neighboring nodes of a particular node are considered.

	1	**2**	**3**	**4**	**5**	**6**
1	0	1	0	0	1	0
2	1	0	1	0	1	0
3	0	1	0	1	0	0
4	0	0	1	0	1	1
5	1	1	0	1	0	0
6	0	0	0	1	0	0

Neighboring nodes of 1: 2, 5.
Neighboring nodes of 2: 1, 3, and 5.
Neighboring nodes of 3: 2, 4.
Neighboring nodes of 4: 3, 5, and 6.
Neighboring nodes of 5: 1, 2, and 4.
Neighboring nodes of 6: 4.

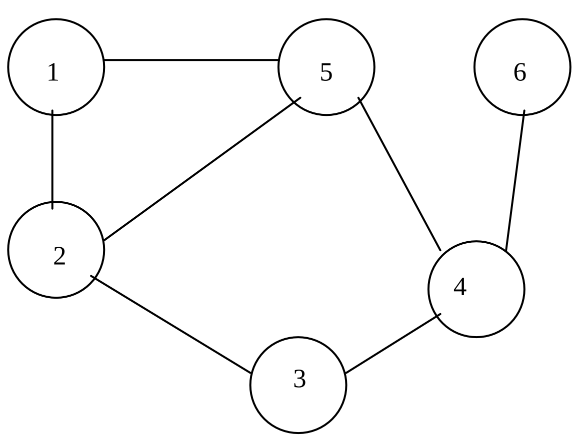


Figure2: Adjacency list and its corresponding matrix

Here, in adjacency list, all the zeroes of the adjacency matrix are eliminated and the wastage of space is also being reduced. But the complexity in computation increases in this case. [3]

This paper concentrates the application and implications of graph theory on social media, social network analysis, interpersonal ties, triadic closure, bridges, and properties of social network. It also includes a case study of few Facebook users, graph partitioning required for structural analysis of social media.

II. Graph Theory and Social Media:

The concept of graph theory is extensively used in social media. Usually here the users or the people involved are considered as the nodes or vertices. And any relation between the users due to common likes or mutual friendship is considered as edges.

- **Graph Theory in Facebook:** Majority are familiar with Facebook these days. You can click 'like' if you find something likeable, 'tag' your friends in various 'posts', put comments in posts and most importantly befriend someone whom you know and also someone whom you don't know! The concept of graph theory is used in Facebook with each person as nodes and every like, share, comment, tag as edges.

- **Graph Theory in Twitter:** Here the persons are considered as nodes and if one person follows another then that is considered as the edge between the two

III. Other Applications of Graph Theory in various fields: [2]

- *Geometrical and other mathematical studies:* If we do not consider the length of the edge and vertical angles of a polygon, then it can be safely assumed as a graph with its vertices and edges. This fact might not be that useful in the study of polygons but this theory is widely used in the study of surfaces and objects with higher dimension. Graph is also used in algebra.

- *Electrical engineering:* Kirchhoff used the concept of graph to understand electrical circuits in the late 1850s and it has been a part of electrical engineering ever since.

- *Designing transportation networks:* Architects and planners use the concept of graph before building highways and bridges. They consider the various cities as nodes and the highway connecting them as edges. It is also used locally inside the cities as well e.g. while constructing a bus stop. Here the bus stops acts as nodes and the roads connecting them are edges. Here the distance between the bus stops i.e. weights or values of the edges are considered because that will indicate the time taken to traverse the distance.

- *Communication networks:* A computer network whether centralized or distributed forms a graph. The internet routing system and other data and packet routing systems in a computer network represents a graphical structure with the various computers or devices as nodes and the routing path between them as edges.
- *The World Wide Web:* The World Wide Web also represents a huge graph. Various pages in the web is considered as nodes and if there is any hyperlink between two pages then that is the edge between those two pages.

IV. SOCIAL NETWORK

To begin with, we need to have some idea about Network. There are several ways of formally defining a network, depending on the branch of mathematics used. The most usual and flexible definition is derived from graph theory, a social network is conceptualized as a graph, that is, a set of vertices (or nodes, units, points) representing social entities or objects and a set of lines representing one or more social relations among them. A network, however, is more than a graph because it contains additional information on the vertices and lines

Formally, a network N can be defined as $N = (U, L, F_U, F_L)$ containing a graph $G = (U, L)$, which is an ordered pair of a unit or vertex set U and a line set L, extended with a function F_U specifying a vector of properties of the units ($f: U \rightarrow X$) and a function F_L specifying a vector of properties of the lines ($f: L \rightarrow Y$). The set of lines L may be regarded as the union of a set of undirected edges E and a set of directed arcs A (L E $\cup$= A). Each element e of E (each edge) is an unordered pair of units u and v (vertices) from U, that is, e(u: v), and each element a of A (each arc) is an ordered pair of units u and v (vertices) from U , that is, a(u: v). [4]
Based on the contents of the nodes Network can be divided into two major types:

- Social and Economic Network – It consists of a group of people connected with some sort of interactions or pattern of communication.
  e.g. - Facebook, Twitter, business relation between companies and clients, interrelationship between families involved in a marriage etc.
- Information Network -The connection between information objects
  e.g. – Semantic (links between various words and symbols), World Wide Web (link between various web pages; new page connecting to another through hyperlinks) [5]

Here, we are going to concentrate only on Social Network and the working principle of Graph Theory on it.

Brief idea on Social Network

When we need to represent any form of relations in the society in the form of links, it can be termed as Social Network. The pattern of interdependency between each individual (node) can be based on different aspects, viz. -friendship, interconnection between families, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige.

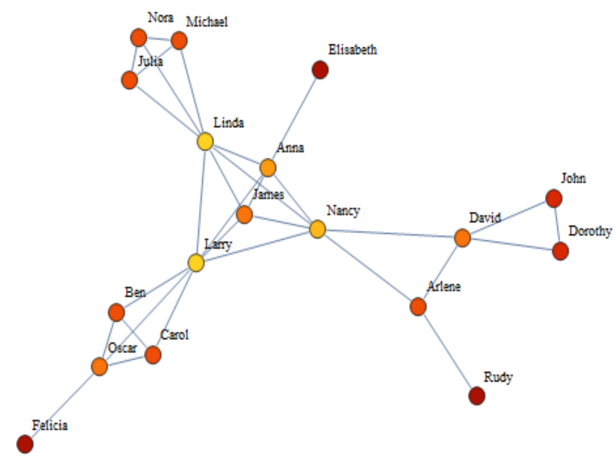


Figure 3: Network of friends

Figure 3 shows the relationship of a particular group of school friends. Each vertex represents a key person in the friendship, and the edges denote a connection by relationship and interaction of any kind that has been denoted by four different colors.

V. SOCIAL NETWORK ANALYSIS

Just like with classical graph theory, there are many points of view from which one can enter the study of social networks. One might come to social networks as a sociologist, or anthropologist, or linguist, or economist, or computer scientist, or engineer, or biologist, or businessperson, or investigator.

The notion of characterizing social networks by key properties is important when it comes to the question of modeling social networks. By modeling social networks, we mean finding an algorithmic way to generate graphs whose key properties are essentially the same as those of social networks found in nature. These methods typically depend upon a random process- maybe at each stage you add/remove a node or edge with some given probability. People have used the idea of **"social network"** loosely for over a century to connote complex sets of relationships between members of social systems at all scales, from interpersonal to international. In 1954, J. A. Barnes started using the term systematically to denote patterns of ties, encompassing concepts traditionally used by the public and

those used by social scientists: bounded groups (e.g., tribes, families) and social categories (e.g., gender, ethnicity). [6]
Overall and Local Network Structure – These are the two perspective to analyze social networks. Overall Network structure concentrates on ties and interaction between persons or other social objects. This approach to social networks is known as the socio-centered approach. The other approach to social networks focuses on the individual element and its immediate network neighborhood and is analyzed through Local Network structure. This is known as the ego-centered approach.

In order to explain and analyze networks, we need to focus on the following topics:

- Interpersonal ties
- Triadic Closure
- Bridges

In mathematical sociology, **interpersonal ties** define the type of connection between two or more people in a relationship. These ties are important and relevant in social network interactions and can be classified into 3 different type based on the strength of interaction: strong ties, weak ties and absent ties.

- Strong ties: the stronger links, corresponding to friends, dependable sources of social or emotional support.
- Weak ties: the weaker links, corresponding to acquaintances.
- Absent ties: the one for which we have no information regarding its nature.

American Sociologist Mark Granovetter refers to your strong ties as your friends and your weak ties as your acquaintances in his paper "Notes on the strength of weak ties" Mark talks about the interpersonal relationships between different, disparate groups of people and how they hold different sections of society together. [7]

Tie Strength in Large-Scale Data- Tie strength refers to a general sense of closeness with another person. Refer to the *Figure 3*, we extend the strong and weak ties as in a continuous quantity to measure the Neighborhood Overlap (*NO*) of an edge *(x,y).*

$$NO(x, y) = \frac{|\text{common neighbors of x and y}|}{|\text{neighbors of at least one of x or y}|}$$

$$= \frac{|N(x) \cap N(y)|}{|(N(x)-\{y\}) \cup (N(y)-\{x\})|}$$

For example, *NO(A, B)* = 0, and *NO(A, F)* = $1/6$ from *Figure 3*.[8]

In "The Strength of weak ties" Granovetter talks about ties and their role in social networks. He defines the strength of a tie as a combination of services, the amount of time, the intimacy, and the emotional intensity between the subjects (Granovetter, 1973). He also talks about the possibility of forming new acquaintances when two persons have a strong connection to a third person, i.e. **triadic closure**. [9]**Triadic Closure** is a principle that implies that if two friends in social network have a friend in common, then there is increased likelihood that they will become friends too. A triad is a group of three entities and triadic closure is the solution to the problem of how an open triad becomes a close triad.

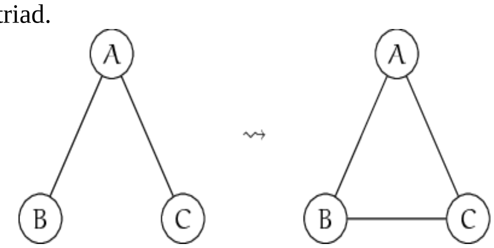


Figure 4: Open triad and close triad

Weak ties play a crucial role in binding a group of strong ties together. They help in strengthening the relationship and forming new bonds between existing relations. So, weak ties can operate as bridges between two groups. To fully understand what this means, one first have to understand the term **bridge**. A link between two nodes A and B is a **bridge** if deleting this link separates the nodes into two components. The bridge is the only way information can move from A to B. There are two types of bridges- local bridge and a regular bridge.

A **local bridge** is a link between two nodes which when broken increases the distance between those nodes to more than two. It's the shortest route through which information can flow from a group of nodes to another group. Based on the definition of *strong triadic closure**, a local bridge is necessarily a weak tie.

The occurrence of a bridge in large social networks is rare due to the fact that in most cases there will be a different path that connects the nodes A and B together. Local bridges however are more often seen and can connect us with parts of the network that would otherwise been neglected. [9]

{**The Strong Triadic Closure Property*–It is the property among three nodes A, B and C, such that if a strong tie exits between A-B and A-C, there is a weak or strong tie between B-C.}

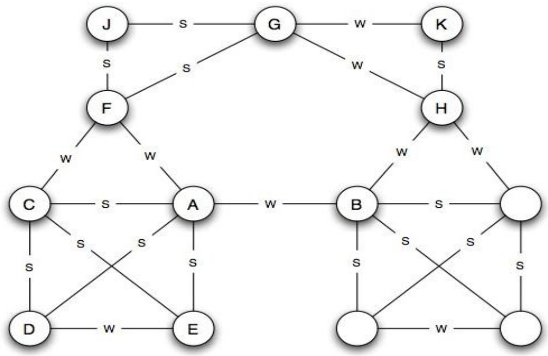


Figure 5: Strong ties, Weak ties, Local bridges

Each edge of the social network is labeled either a strong tie or weak tie to indicate the strength of a relationship. Local bridges are especially those with large span.

Regarding absent ties, Rosenfeld and Thomas (2012) argue that "the Internet increasingly allows Americans to meet and form relationships with perfect strangers, that is, people with whom they had no previous social tie". [10]

Structural holes were first developed by Ronald Stuart Burt in 1992. It is concept from social network. It is a gap between two individuals who have complementary sources to information. It spans the fields of sociology, economics, and computer science. [11]Structural hole is lack of connection between two nodes that is bridged by a broker.

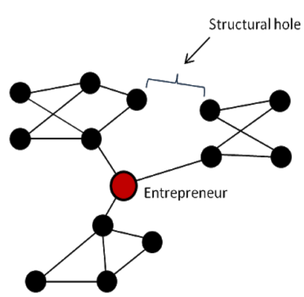


Figure 6: Structural Hole

VI. Properties of Network

Euler unintentionally showed that networks have properties, which is dependent on their topological architecture that limit or enhance their usability. Network analysis studies properties at three levels of abstraction:

- **Element-level analysis**, where methods to identify the most important nodes of the network are investigated,
- **Group-level analysis**, that involves methods for defining and finding cohesive groups of nodes in the network, and
- **Network-level analysis**, that focuses on topological properties of networks as a whole. [12]

Small networks can be visualized by their graphs but larger networks can be difficult to be analyzed. Therefore, we define a set of summary statistics or quantitative performance measures to describe and compare networks (focus on undirected graphs):

- Size
- Density
- Planer Network Density
- Connectedness
- Diameter and average path length
- Clustering
- Centrality
- Degree distributions

Here we will explain some of the characteristics:

- **Diameter and average path length**-Let I (i, j) denote the length of the shortest path (or geodesic) between node i and j (or the distance between i and j). The diameter of a network is the largest distance between any two nodes in the network:

Diameter= max I (i, j)

The average path length is the average distance between any two nodes in the network:

Average path length = $\dfrac{\sum_{i \geq j} I\ (i, j)}{\frac{n(n-1)}{2}}$

- ✓ Average path length is bounded from above by the diameter; in some cases, it can be much shorter than the diameter.
- ✓ If the network is not connected, one often checks the diameter and the average path length in the largest component.

- **Clustering**-We can measure the strength of triadic closure via the clustering coefficient. This clustering measure is represented by the overall clustering coefficient Cl(g), given by

Cl(g)= $\dfrac{3 \text{ X } number\ of\ triangles\ in\ the\ network}{number\ of\ connected\ triples\ of\ nodes}$

Where a "connected triple" refers to a node with edges to an unordered pair of nodes.
Note that $0 \leq Cl(g) \leq 1$.

Cl(g) measures the fraction of triples that have their third edge filled in to complete the triangle.

Another measure of clustering is defined on an individual node basis:

The individual clustering for a node ***i*** is

$$Cl_i(g) = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered at } i}$$

The average clustering coefficient is $Cl^{Avg}(g) = 1/n \sum_i Cl_i(g)$

- **Centrality**- A micro measure that captures the importance of a node's position in the network.

Different measures of centrality:

- *Degree centrality*: for node i , $d_i(g)/n - 1$, where di (g) is the degree of node i

- *Closeness centrality*: Tracks how close a given node is to any other node: for node i , one such measure is

$$\frac{n-1}{\sum_{j \neq i} l(i,j)}$$, where *l (i,j)* is the distance between i and j

- *Betweenness centrality*: Captures how well situated a node is in terms of paths that it lies on.

Centralization is the corresponding structural property of the overall network and it is defined as the variation in the centrality scores of the vertices in the network because this variation shows the extent to which there is a center (very central vertices) and a periphery (vertices with very low centrality scores). The star and ring networks are defined as respectively the most and least centralized networks (centralization scores of 1 and 0).

- **Degree distributions**- The degree distribution, P(d), of a network is a description of relative frequencies of nodes that have different degrees d.
  - For a given graph: P(d) is a histogram, i.e., P(d) is the fraction of nodes with degree d.
  - For a random graph model: P(d) is a probability distribution. [5]

VII. Network Models

Network models can be classified based on two factors:

1. Topology- Some common models are:
   - Random graph
   - Erdős–Rényi
   - Barabási–Albert
   - Fitness model
2. Dynamics- Known models are given below
   - Boolean Network
   - Agent Based
   - Epidemic/SIR

VIII. Example of Case Study: Measuring Networks on Facebook [13]

This case study depicts the size of Facebook networks. Few random users have been studies for the course of a month and the relationship is classified into four different network patterns:

All Friends: This network represents the list of all friends a user has, hence this is the largest among all the representations

Reciprocal Communication: This representation shows the mutual communication between two parties, this type of network forms when there is mutual exchange of information between two parties.

One-way Communication: It consists of people with whom a user has communicated.

Maintained Relationships: This pattern of relationship consists of people whose profile has been checked by the user more than once to maintain engagement.

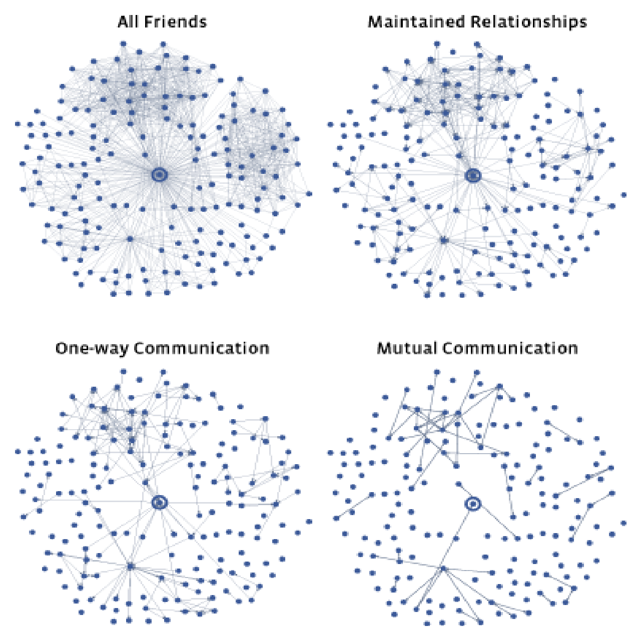


Figure 7: Network patterns of a user on Facebook

In the diagram, the red line shows the number of reciprocal relationships, the green line shows the one-way relationships, and the blue line shows the passive relationships is a function of your network size.

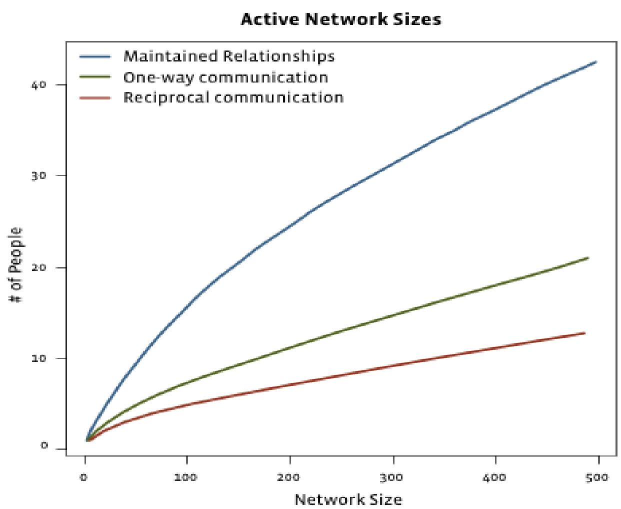


Figure 8: Graphical representation

IX. GRAPH PARTITIONING

Graphs are also used by various scientists or researchers for modeling an application program. Partitioning a graph is purely an algorithmic problem. It helps to reduce complexity of big graphs and also introduces parallelization. Graph partitioning is required in various application problems like social networks, road networks, air traffic controls, image analysis etc. Some applications of graph portioning problems are scientific computing, partitioning various stages of VLSI design circuit, task scheduling in multiprocessor systems etc. The aim of graph partitioning is to divide the nodes into several disjoint parts such that the predefined objective function is minimal. The optimal graph partitioning is NP-complete however various approximate algorithms are made to solve the problems.

Graph partitioning is divided into two groups –

- **i) Constrained Partitioning**- In this partitioning the parts are of equal size.
- **ii) Unconstrained Partitioning**- In this portioning the parts are of different size.

Various algorithms were developed for graph partitioning. Among them three principle algorithms are Geometric partitioning, Spectral partitioning and Multi-level graph partitioning. [14]

In Geometric partitioning the graph is bisected by utilizing those coordinates which are obtained if nodes of a graph are available in space. In this partition the vertices which are spatially near to each other are taken into one cluster.

In Spectral partitioning the associability of the graph is concluded by finding the Eigen vectors with respect to the second smallest Eigen value of Laplacian matrix *L* corresponding to graph. This bisection method is really demanding but not feasible for large graphs.

Multi-level partitioning is highly effective than the classical graph partitioning methods. Constrained graph partitioning problems are efficiently solved by this method. The main idea is to partition the large graphs into *k* – parts, group the vertices together in a group and deal with this group of vertices rather than independent vertices. It has three phases – coarsening phase, initial partitioning phase, and partition refinement phase. [14]

X. STRUCTURAL ANALYSIS

Graph is a mathematical structure which shows relation between some objects. Graph is made up of vertices, nodes, or points which are connected using lines, arcs or edges. A graph may directed or undirected. A graph represented as $G= (V, E)$ where V is set of vertices and E is set of edges. Graph is represented using different data structures like adjacency list, adjacency matrices, incidence matrices etc.

In a social network the node represent individuals or organizations and the edges represent the relationship between individuals or organizations. Social network provides set of methods for analyzing the structure of whole social entities. There are different levels of analysis that are not necessarily mutually exclusive. Generally there are three levels:

- *Micro level*- At this level it typically begins with an individual or may begin with a small group of individuals.
- *Meso level*- It falls between micro and macro-level. It shows the connection between the micro and macro level. Meso-level networks are low density network.
- *Macro level*- These are large scale networks that traces the interactions over large population. These networks are more complex.

XI. CONCLUSION

In this paper we have discussed about the various aspects of graphs and its basic properties. Graphs are used in social networks whose complexity is increasing with the advancement in social media. Researchers are still developing various algorithms for better understanding of social networks and for partitioning the graphs to reduce its complexity and for better understanding of networks.

REFERENCES

[1]. Marcin Mince, Ewa Niewiadomska-Szynkiewicz, "Application of Social Network Analysis to the Investigation of Interpersonal Connections", February 2012.

[2]. Kimball Martin, "Graph Theory and Social Networks", April 30, 2014.

[3]. Alexandru Costan, "Graph Theory and Social Networks",

[4]. 4.https://www.uva.nl/binaries/content/documents/personalpages/n/o/w.denooy/en/tab-one/tab-one/cpitem%5B26%5D/asset?1355372751494.
[5]. Daron Acemoglu and Asu Ozdaglar, "Networks, Lecture 2: Graph Theory and Social Networks", September 14, 2009.
[6]. Jayson Rome, "Graph Triangulation", October 14, 2002.
[7]. Eileen Brown, "SOCIAL MEDIA TODAY: Strong and Weak Ties: Why Your Weak Ties Matter", June 30, 2011.
[8]. Donglei Du, "Strength of weak ties paradox".
[9]. Sosial nettverksteori in blog tilhorende kurset INFO207/INF207 pa Universitetet i Bergen, innlegg skrevet av studentene med'
[10]. Josu´e Ortega, Philipp Hergovich, "The Strength of Absent Ties: Social Integration via Online Dating", October 2, 2017.
[11]. https://www.semanticscholar.org/topic/Structural-holes/448117
[12]. 12.https://users.dimi.uniud.it/~massimo.franceschet/networks/nexus/properties.html
[13]. Cameron Marlow, Lee Byron, Tom Lento, Itamar Rosenn, "Maintained relationships on Facebook", 2009.
[14]. Tieyun Qian1, Yang Yang2, Shuo Wang2, "Refining Graph partitioning for Social Network Clustering",2010.

Authors Profile

Anwesha Chakraborty is currently pursuing M.Sc. in Computer Science (2017-2019) from St. Xavier's College (Autonomous) Kolkata. She has received her B.Sc. degree in Computer Science in 2017 from the same college. Her final year project on B.Sc. was based upon Virtual Classroom, a web based project. She has also done a review paper on Li-Fi, Big Data Analytics and on Java, defining it with the help of Finite state automation.

Trina Dutta is currently pursuing M.Sc. in Computer Science (2017-2019) from St. Xavier's College (Autonomous) Kolkata. She has received her B.Sc. degree in Computer Science in 2017 from the same college. Her final year project on B.Sc. was based upon Virtual Classroom, a web based project. She has also done a review paper on Li-Fi, and on Big Data Analytics.

Sushmita Mondal is currently pursuing M.Sc. in Computer Science (2017-2019) from St. Xavier's College (Autonomous) Kolkata. She has received her B.Sc. degree in Computer Science in 2017 from the same college. Her final year project on B.Sc. was based upon Virtual Classroom, a web based project. She has also done a review paper on Li-Fi, and on Big Data Analytics.

Dr. Asoke Nath is working as lecturer in the Department of Computer Science, St. Xavier's College (Autonomous), Kolkata. He is engaged in research work in the field of Cryptography and Network Security, Steganography, Green Computing, Big data analytics, Li-Fi Technology, Mathematical modelling of Social Area Networks, MOOCs etc. He has published more than **232** research articles in different Journals and conference proceedings

