2. COMMUNITY DETECTION IN SOCIAL NETWORK

The modern networks like the social network can generally be modeled as a graph structure and the problems like link prediction, community detection in social network analysis and social network mining can be solved using graph theory techniques. One of the most relevant features of graphs representing real systems is community structure or cluster i.e. the organization of vertices, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. Such clusters, or communities, can be measured as fairly independent compartments of a graph, playing a similar role. Detecting communities is of great importance in disciplines where systems are often represented as graphs. This chapter presents the basic concepts of graph theory and theoretical background of community detection.

2.1 GRAPH THEORY - DEFINITION AND TERMINOLOGY

Most of the social networks are directed graphs which is a graph with a set of nodes connected by edges, where the edges have a direction associated with them. This section provides the basic terminology and graph theory background that is used throughout the research work.

Graph

A network is usually represented by a graph. A graph G = (V, E) consists of a set of nodes V and a set of edges E ⊆ V × V which connect pairs of nodes. The number of nodes in the graph is equal to n = |V | and the number of edges m = |E|. Table III gives a list of symbols used along with their definition.

Table III Symbols and Definitions

Symbol	Definition
G	Directed network
G_U	Undirected network
$G_B = (V_h, V_a, E_b)$	Bipartite network
V, E	Set of nodes and edges for network G
$\|V\| = n, \|E\| = m$	Number of nodes and edges in the network
e = (u, v)	Edge e ∈ E from node u to node v
A_U, A	Adjacency matrix of an undirected and directed network respectively

k_u^{in}, k_u^{out}	In degree and Out
D_{in}, D_{out}	Diagonal In and Out
A_{ij}	Entry of matrix A
A^T	Transpose of matrix A
λ_i	i^{th} largest eigenvalue of a matrix
u_i	Eigenvector corresponds to i-th eigenvalue
u_{ij}	i-th component of j-th eigenvector

Nodes

All graphs have fundamental building blocks. One major component of any graph is the set of nodes. In a graph representing friendship, these nodes represent people, and any couple of connected people denotes the friendship between them. Depending on the perspective, these nodes are called vertices or actors. For example, in a web graph, nodes represent websites, and the connections between nodes indicate web-links between them. In a social setting, these nodes are called actors. The mathematical representation for a set of nodes is $V = \{v_1, v_2, \ldots v_n\}$ where V is the set of nodes and v_i, $1 \leq i \leq n$, is a single node. $|V| = n$ is called the size of the graph.

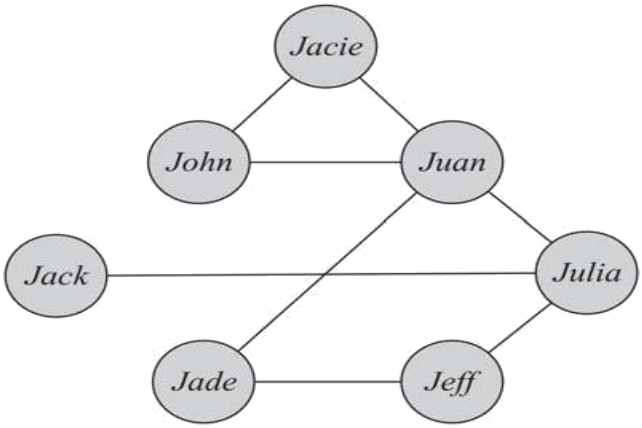
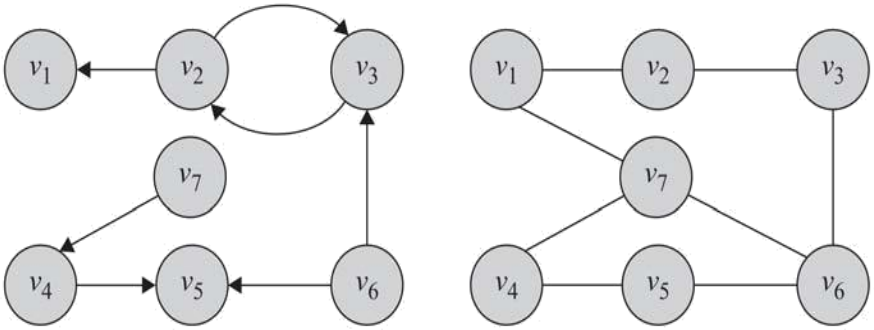


Fig. 2.1 A Sample Graph

Edges

Another important component of any graph is the set of edges. Edges connect nodes. In a social setting, where nodes represent social entities such as people, edges indicate inter-node relationships and are therefore known as relationships or ties. The edge set is usually

represented as E, such that E = {e_1,e_2,… e_m} where e_i, 1≤ i≤ m, is an edge and the size of the set is commonly shown as m = |E|. In Fig 2.1 lines connecting the nodes represent the edges.



(a) Directed Graph (b) Undirected Graph

Fig. 2.2 A Directed Graph and an Undirected Graph

Edges are also represented by their endpoints, so e (v_1; v_2) defines an edge e between nodes V_1 and V_2. Edges can contain directions, meaning one node is linked to another, but not vice versa. When edges are undirected, nodes are linked both ways. In Fig.2.2 (b), edges e (v_1, v_2) and e (v_2, v_1) are the same edges, because there is no direction stating how nodes get connected. These edges in this graph are called undirected edges and this kind of graph an undirected graph. Conversely, when edges include directions, e (v_1, v_2), is not the same as e (v_2, v_1).

A graph can be directed or undirected, unipartite or bipartite and the edges may contain weights or not. Graph shown in Fig. 2.2(a) is a graph with directed edges, an example of a directed graph. Directed edges are represented using arrows. In a directed graph, an edge e (v_i, v_j) is characterized using an arrow that starts at v_i and ends at v_j. Edges can begin and end at the same node; these edges are called loops or self-links and are represented as e (v_i, v_i). For any node v_i, in an undirected graph, the set of nodes connected to via an edge is called its neighborhood and is represented as neighborhood N(v_i). In Fig 2.1, N (Jade) = {Jeff, Juan}. In directed graphs, node v_i has arriving neighbors N_{in} (v_i) (nodes that connect to v_i) and outgoing neighbors N_{out}(v_i). In Fig 2.2(a), N_{in} (v_2) = {v_3} and N_{out}(v_2) = {v_1, v_3}.

The mathematical definitions are given below.

Directed and Undirected Graph: In a directed graph G = (V, E), every edge (i, j) ∈ E links node i to node j. An undirected graph G_U = (V, E) is a directed one where if edge (i, j) ∈ E, then edge (j, i) ∈ E.

Bipartite Graph: A graph $G_B = (V_h, V_a, E_b)$ is called bipartite if the node set V can be partition into two dislodge sets V_h and V_a, where $V = V_h \cup V_a$, such that every edge $e \in E_b$ connects a node of V_h to a node of Va, i.e., $e = (i, j) \in E \Rightarrow i \in V_h$ and $j \in V_a$. In other words, there are no edges between nodes of the similar partition.

Adjacency Matrix: Every graph G = (V, E) directed or undirected, weighted or unweighted can be represented by its adjacency matrix A. Matrix A has size $|V| \times |V|$ (or $n \times n$), where the rows and columns represent the nodes of the graph and the entries indicate the existence of edges.

The adjacency matrix A of a graph G = (V, E) is an $|V| \times |V|$ matrix, such that

$$A_{ij} = \{w_{ij}, \text{if } (i,j) \in E, \forall i,j \in 1,\ldots,|V| \quad (2.1)$$

$$0, \text{otherwise}$$

This definition is suitable both for weighted and unweighted graphs. For the former case, each value w_{ij} represents the weight associated with the edge (i, j), while for the latter case of unweighted graphs the weight of each edge is equal to one (i.e., $w_{ij} = 1, \forall (i, j) \in E$). If the graph is undirected, the adjacency matrix A is symmetric, i.e., $A = A^T$, while for directed graphs the adjacency matrix is non-symmetric.

Degree: A basic property of the nodes in a graph is their degree. In an undirected graph G_U, nodes have degree k if it has k incident edges. In case of directed graphs, every node is associated with an in-degree and an out degree. The in-degree k in i of node $i \in V$ is equal to the number of incoming edges, i.e., k in i = kj| (j, i) $\in E_k$, while the out-degree k out i of node $i \in V$ equals to the number of outgoing edges, i.e., k out i = k_j|(i, j) $\in E_k$. In undirected graphs, the in-degree is equal to the out-degree, i.e., k_i = k in i = k out i, $\forall i \in V$. The degree matrix is defined as the diagonal $n \times n$ matrix D, with the degree of each node in the main diagonal. Similarly, in directed graphs the in-degree matrix D_{in} and out-degree matrix D_{out} for the in- and out- degrees can be defined [61].

Terminology

Edge betweenness: The betweenness of an edge is the number of shortest paths between vertices that contain the edge.

Random walks: **It** is a Markov chain which describes the sequence of nodes visited by a random walker. Random walk is used to calculate the dissimilarity between two nodes in order to identify community.

Centrality: It is a measure indicating the importance of node in the network.

Degree Centrality: It is defined as the ratio of the number of neighbours of a vertex with the total number of neighbours possible.

In Degree: This represents the number of edges incoming to a vertex.

Out degree: This represents the number of edges outgoing from a vertex.

Clique: A clique is a maximum complete sub graph in which all nodes are adjacent to each other.

Maximum clique: A maximum clique is the largest clique in a graph, a subset in which all vertices are pair-wise connected by an edge.

Complete mutuality: It is a measure of tie strength inside the subgroup.

Reachability: It is a low diameter, facilitating fast communication between the group members.

k-clique: *It is* a maximal sub graph in which the largest geodesic distance between any two nodes is no greater than k.

k-core: The k-core is defined as the largest sub graph in which each node has at least k edges. The k-core graph was used to find the maximal sub graph with minimum degree k. k-core is a substructure that each node connects to at least k members within the group.

k-plex: k-plex of a graph is a maximal subgraph in which each vertex of the induced subgraph is connected to at least n-k other vertices, where n is the number of vertices in the induced subgraph.

Modularity: Modularity measures the excellence of community partitions formed by an algorithm. It is the dissimilarity between the actual density of intra-community edges and the corresponding connections in a random network possessing the same degree distribution as that of the actual network.

Overlapping: Overlapping communities are possible if a node is a member of more than one community.

2.2 COMMUNITY AND ITS TYPES

A community is a set of entities where each entity is closer to other entities within the community than to the entities outside it. One of the widely-used definitions of communities is based on the number of edges within a group compared to the number of edges between different groups. A community is thought of as a group of nodes that has more well-connected edges between its members than the remainder of the network. A community helps to understand the structure of social networks because communities are considered as

components of social networks and identify the features as well as roles of the network. Communities facilitate the visualization of large-scale social networks. Communities enable the process of information sharing and dissemination of information to the members through the network.

Communities can be implicit or explicit. Communities that are not actually built by its group members but formed by a third party come under implicit category. For example, yahoo groups come under explicit community, whereas a community in which all the people who use similar or same programming languages come under implicit. In most of the social networking sites, contrast to explicit communities, implicit communities and their members are obscure to many people.

Community structure is defined as the possibility of recognizing within the networks, subsets of nodes which are more connected among themselves that to the rest of the network. When detecting communities, there are two possible modes of data (1) the network structure, (2) the features and attributes of nodes. Even though communities form around nodes that have common edges and common attributes, community detection algorithms have only focused any one of these two data modalities. Traditional community detection algorithms concentrate only on the network structure, while clustering algorithms mostly consider only node attributes [62].

Nature of Network Communities

As the complexity of networks increases, the definitions of community also differ. Though there is variety of different definitions of community, there are many features to be considered in the problem of community identification from complex networks. In this section, some of the important natures of communities are discussed.

Directed Network: Some real-world networks are represented with edges and links, that are not reciprocal called directed network. For example, in case of web pages, a hyperlink from one page to another is directed and other page may or may not have a hyperlink pointing in the backward direction. In community detection, direction of edges also plays an important role.

Hierarchy Network*:* When a node in a network belongs to only one community, then the community is said to be separated or disjoint. Most of the disjoint communities are hierarchical in nature. Hierarchy describes the organization of elements in a network. It shows how nodes link to each other to form motifs, how motifs combine to form communities and how communities are joined to form the entire network. In general, a

network's community structure encompasses a potentially complicated set of hierarchical and modular components. In this context, the term module is used to refer to a single cluster of nodes. Given a network that has been partitioned into modules, it can be divided in an iterative fashion until each node is in its own singleton community [63].

Overlapping Network: The complex network models of real-world phenomena exhibit an overlapping community structure, i.e. a node in the network can belong to more than one community. The presence of nodes belonging to several communities occurs naturally from real data. Hence, overlap is one of the peculiar features of community. The overlap of different communities exists widely in real-world complex networks, particularly in social networks. In complex networks, nodes are typically shared between two or more groups. In such cases, communities are said to be overlapping. Fig.2.3 shows the disjoint community structure and Fig.2.4 shows an example of possible overlapping of nodes by two communities.

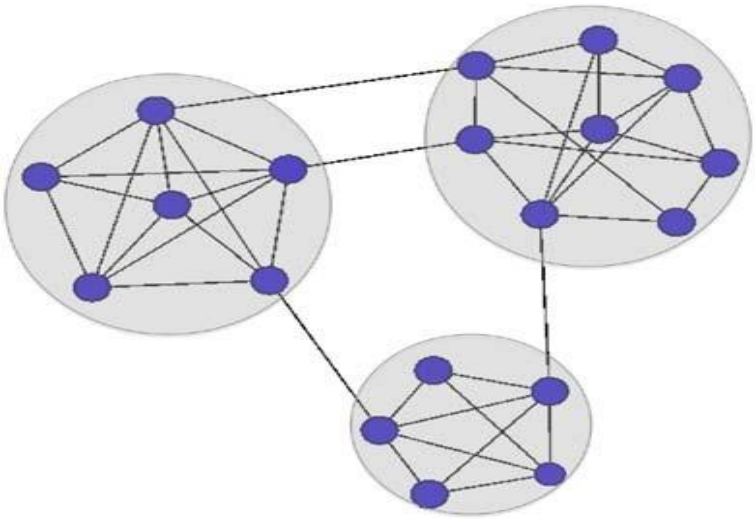


Fig. 2.3 Disjoint Communities

Dense groups in complex networks often overlap with each other. For example, in social networks, human beings have multiple roles in the society and these roles make the members of network to join into multiple communities at the same time such as colleges, universities, families or relationships, companies, hobby clubs, etc. In co-authorship network, nodes represent the scientists and two nodes are connected if they have coauthored one or more articles and the articles are communities. Overlapping considerably increases the complexity of the communities [64].

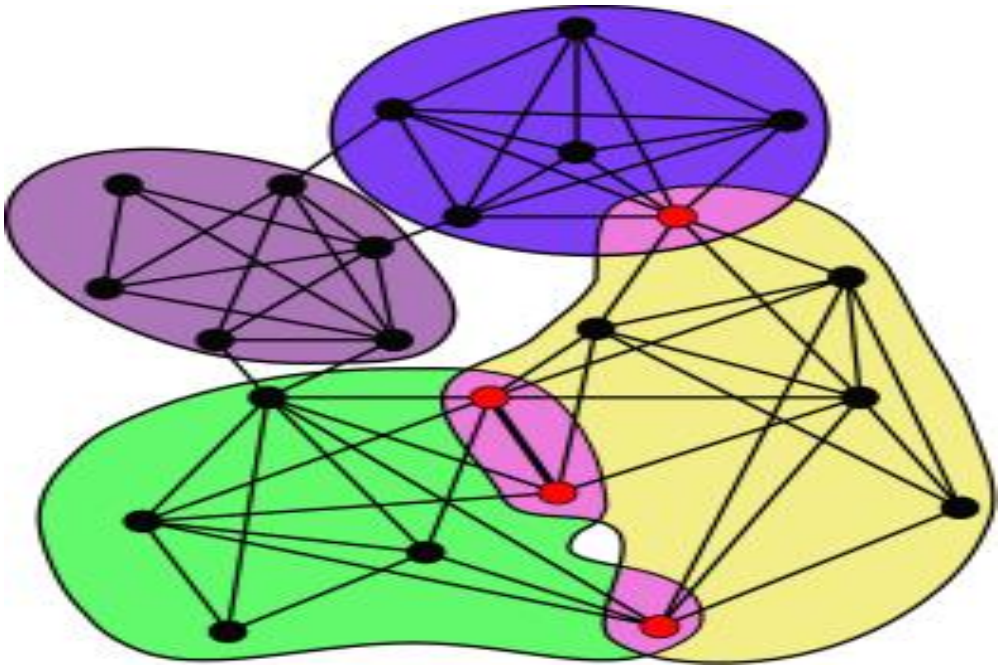


Fig. 2.4 Overlapping Communities

Weighted Network: A weighted network is a network where the links among nodes have weights assigned to them. In many real-world networks, not all links in a network have the same capacity. Links in complex networks are often associated with weights that differentiate them in terms of their strength, intensity or capacity. In social networks, the strength of social relationships is a function of their duration, emotional intensity, intimacy and exchange of services. For non-social networks, weights often refer to the function performed by links. In weighted networks, a group of vertices can be considered as a community only if the weights of their connections are strong enough [65].

Dynamic Network: Complex networks are not always static. In reality, networks gradually evolve over time. Particularly, social networks witness the expansion in size and space as their users continuously increase, changing the network to dynamic in nature. A dynamic network is a special type of evolving complex network where changes are often introduced over time. The set of edges appearing and disappearing in the communities as time evolves have a little effect to the local structure of the network. But, over a long period of time this dynamics may lead to a significant transformation of network community structure. The study of dynamic communities is an emerging area of interest in the field of complex networks [66].

2.3 COMMUNITY DETECTION APPROACHES

The actors in a network tend to form groups of closely-knit connections. The groups are also called communities, clusters, cohesive subgroups or modules in different perspective. Individuals interact more frequently within a group than between groups. Detecting cohesive groups in a social network remains a core problem in social network analysis. Finding out these groups also aids for other related tasks of social network analysis. Classic approaches of finding communities in network borrow the design of graph partitioning and hierarchical clustering. Graph partitioning approaches needs to recognize information about the global structure of network and determine in advance the number and size of subgroup that they want to get. Hierarchical partitioning is cluster analysis method in which the network of interest is divided into several subgroups. The division is natural because it depends on node relationship inside the network than node properties itself. Node relationship is measured by similarity metrics, such as vertex similarity and edge betweeness. Both metrics uses corresponding matrix and has the drawbacks on computation complexity, when it come to large-scale network. From many different ideas and perspective, the community detection based research roughly categorized into four approaches Node-Centric, Group-Centric, Network-Centric, and Hierarchical-Centric [67]. The detailed note on these approaches is presented below.

2.3.1 Node Centric Community Detection

Node centric community detection approaches generally determines groups where each node in a group satisfy important properties such as complete mutuality, reachability, nodal degree, relative frequency. Complete mutuality of nodes in a group defines a clique which is a fully connected subgroup. Maximum clique approach is a basic algorithm for community detection. Clique percolation approach is an enhanced version of maximal clique which is used for finding overlapping communities. Reachability among nodes happens if there exists a path between those nodes. This property of nodes defines k-clique which is a maximal subgraph wherein the largest geodesic distance between any of two nodes is no greater than k. k-clique is commonly used in traditional SNA [67]. Detailed descriptions of some of the node-centric community detection algorithms are presented in section 2.4.

2.3.2 Group-Centric Community Detection

In group centric community detection each node in the group has to satisfy certain properties. Group-centric criteria consider the connections inside a group as whole. It is acceptable to have some nodes in the group to have low connectivity as long as the group

satisfies certain requirements. One such example is density-based groups. Density based group does not guarantee the nodal degree or reachability for each node in the group. It allows the degree of different nodes to vary drastically, thus seems more suitable for large-scale networks.

A greedy algorithm is adopted to find a maximal quasi-clique. The quasi-clique is initialized with a vertex with the largest degree in the network, and then expanded with nodes that are likely to contribute to a large quasi-clique. This expansion continues until no nodes are added to maintain density. Due to the heuristic being used, not all satisfied communities can be enumerated. But it is able to identify some communities for a medium range of community size/density, to detect small communities [67].

2.3.3 Hierarchy-Centric Community Detection

Another line of community detection is to build a hierarchical structure of communities based on network topology. This facilitates the examination of communities at different granularity. There are mainly three types of hierarchical clustering: divisive, agglomerative, and structure search.

Divisive hierarchical clustering: Divisive clustering first partitions the actors into several disjoint sets. Then each set is further divided into smaller ones until the set contains only a small number of actors. Here the key is how to split the network into several parts. Some partition methods can be applied recursively to divide a community into smaller sets. One particular divisive clustering proposed for graphs is based on edge betweeness. It progressively removes edges that are likely to be bridges between communities. If two communities are joined by only a few cross-group edges, then all paths through the network from nodes in one community to the other community have to pass along one of these edges. Edge betweenness is a measure to count how many shortest paths between pair of nodes pass along the edge, and this number is expected to be large for those between-group edges. Hence, progressively removing those edges with high betweenness can gradually disconnect the communities, which leads naturally to a hierarchical community structure [67].

Agglomerative hierarchical clustering: Agglomerative clustering begins with each node as a separate community and merges them successively into larger communities. Modularity is used as a criterion to perform hierarchical clustering. A community pair is merged such that it results in the largest increase of overall modularity and the merge continues until no merge can be found to improve the modularity. This algorithm incurs many imbalanced merges, resulting in high computational cost. Hence, the merge criterion is modified accordingly to

take into consideration the size of communities. In such scheme, communities of comparable sizes are joined first, leading to a more balanced hierarchical structure of communities and to improved efficiency.

Structure Search: Structure search starts from a hierarchy and then searches for hierarchies that are more likely to generate the network. This idea was first implemented in topic taxonomy for group profiling, and then a similar idea is applied for hierarchical construction of communities in social networks. It defines a random graph model for hierarchies such that two actors are connected based on the interaction probability of their least common ancestor node in the hierarchy. A sequence of hierarchies is generated via local changes of the network based on the probability of likelihood. The final hierarchy is the consensus of a set of comparable hierarchies. The bottleneck with structure search approach is its huge search space. The challenge is how to scale it to large networks [67].

2.3.4 Network-Centric Community Detection

The network-based community detection algorithms are based on the overall topology of the network, aiming at obtaining possible partitions from within the network. These algorithms usually include certain metrics defined upon all the partitions. Some of the algorithm types that follow the network-centric approach are given below.

***Vertex Similarity*:** Vertex similarity community detection algorithm is based upon a similarity metric between pair of nodes. A commonly used similarity is the structural similarity, where the similarity is defined by how much any pair of nodes connects to similar nodes. Once this similarity measures are obtained for all nodes in the system, any clustering algorithm can be used to obtain the different groups. A sample similarity measures include Jaccard and Cosine similarity measures.

Modularity Maximizations: Modularity is a measure of the quality of the network partitions. Modularity as a metric and modularity maximization as a community detection algorithm are of special importance. A number of community detection algorithms use the principle of modularity maximization to find the most optimal structure. A possible approach here is through a greedy algorithm. It begins off with all nodes belonging to their own separate communities and then merges those two communities that make a better overall modularity score. The process continues until a local modularity maximum is found. Another important algorithm that utilizes modularity is Girvan - Newman algorithm [67].

This research work employs node-centric approach for community detection in social network and is elucidated in the following section.

2.4 NODE-CENTRIC COMMUNITY DETECTION ALGORITHMS

A community is a densely connected subset of nodes that is sparsely linked to the remaining network. Social networks are a combination of important heterogeneities in complex networks, such as collaboration networks and interaction networks. Finding communities within an arbitrary network is an interesting and computationally difficult task. Node-centric community detection algorithms are basically grouped into (i) principal community detection algorithms (ii) subgraph community detection algorithms (iii) overlapping community detection algorithms.

Principal Community Detection Algorithms

Girvan-Newman and Random Walks algorithms are the basic community detection techniques designed for directed graph.

***Girvan-Newman Partitioning Algorithm*:** The Girvan-Newman technique for the detection and analysis of community structure depends upon the iterative elimination of edges with the highest number of the shortest paths that pass through them. By getting rid of the edges, the network splits down into smaller networks, i.e. communities. The idea is to find which edges in a network occur most frequently between other pairs of nodes by finding edge betweenness. The edges union communities are then expected to have high edge betweenness. The underlying community structure of the network will be fine-grained once edges with high edge betweenness are eliminated.

Girvan-Newman algorithm follow the below steps:

Step 1: Calculate edge betweenness for every edge in the graph

Step 2: Remove the edge with highest edge betweenness

Step 3: Calculate edge betweenness for remaining edges

Step 4: Repeat steps 2–4 until all edges are removed.

In order to calculate edge betweenness, it is necessary to find all shortest paths in the graph. The algorithms begin with one vertex, calculates edge weights for paths going through that vertex, and then repeats it for every vertex in the graph and amount the weights for every edge [68].

Random Walks Algorithm: Random walks algorithm is another commonly used community detection algorithm designed for directed graph. Let G be a graph or digraph with the additional assumption that if G is a digraph, then $\deg^+(v) > 0$ for every vertex v. Consider an object placed at vertex v_j. At each stage the object moves to an adjacent vertex. The

probability that it moves to the vertex v_i is $\frac{1}{\deg(vi)}$ or $\frac{1}{deg+(vj)}$ if (v_j, v_i) is an edge on G and G is a graph or digraph, respectively. Otherwise the probability is 0. Therefore

$$mij = \begin{cases} \frac{1}{\deg(vj)} \; if \; (vj, vi) is \; edge \; in \; the \; graph \; G \\ \frac{1}{deg + (vj)} \; if \; (vj, vi) is \; an \; edge \; in \; the \; digraph \; G \\ 0 \; otherwise \end{cases} \quad (2.3)$$

Then M = (m_{ij}) is a Markov matrix. The roles of i and j are reversed as it need the columns of M to sum to 1. As each stage, a sequence of adjacent vertices is produced. This sequence signifies the 1 position of the object at a given stage. Moreover, this sequence is a walk in the graph. It is called such a walk as a random walk on the graph or digraph G. The i, j entry of M_k represents the probability that a random walk of length k starting at vertex v_j, ends at the vertex v_i [69].

Sub-Community Detection Algorithms

A subgraph of graph G is a graph G' such that G' ⊆ G thus possessing a smaller set of the vertices and edges of the parent graph. An induced sub graph is a subgraph G' of a graph G, where all edges connecting the vertices V' in G' are also present in G. An edge-induced subgraph is a set of edges taken from the parent graph, in which vertices incident to the edges are included [56]. A subgraph is a common subgraph of graphs G_1 and G_2 if it is isomorphic to the subgraphs G'_1 and G'_2 of G_1 and G_2 respectively. The Maximum Common Induced Subgraph (MCIS) is the largest induced subgraph common to G_1 and G_2, whereas the Maximum Common Edge Subgraph (MCES), is the largest number of edges isomorphic to G_1 and G_2. There are three types of subgraphs namely Maximal k-clique, maximal k-core, maximal k-plex. Most of the sub-community detection algorithms are based on these three types of subgraphs [70].

Maximal k-clique: A clique in graph theory is a series of vertices such that each vertex is connected to each other vertex. A maximal clique in a graph is a clique that cannot be extended i.e. no nodes can be added to it which can enlarge the clique. A maximum clique of a graph is a maximal clique of the largest possible size, bringing the possibility that a graph can in fact possess multiple maximum cliques. Clique detection algorithms make up a group of methods which are important for finding the Maximal Clique Subgraph (MCS) of graphs. The maximum clique problem is to find a largest possible complete sub-graph. A triangle is an example of a clique, where all the nodes are connected to each other. Maximum clique

algorithms are being widely used within computing science in areas like, computer vision, design of communication protocols, compiler code generation, malware detection, cryptography, robotics, fraud detection, fault diagnosis, manufacturing, and sociology [71].

Maximal k-Core: A *k*-core of a graph is a maximal connected subgraph in which every vertex is connected to at least *k* vertices in the subgraph. A k-core framework consists of three steps and uses any one of standard community detection methods in an inner step. The first step is to reduce the whole graph to a k-core. The second step uses an existing algorithm to generate community labels for nodes in the k-core. The third step is to find community labels for the remainder of the graph via a fast algorithm. The pseudocode given below is the k-core framework where, G is the original graph, K is the desired core number, G_K is the k-core subgraph, G_K is the community assignment for the k-core, and G is the community assignment for all nodes [72].

Algorithm to find the k-core Subgraph

Step 1: input Graph G, Parameter K

Step 2: output Subgraph G_K

Step 3: G_K G

Step 4: while G_K is not a k-core do

Step 5: Find all nodes in G_K whose degree is less than K

Step 6: Remove those nodes and their incident edges

Step 7: Update the node degrees for the remaining nodes

Step 8: end while

Step 9: return G_K

Maximal k-plex: A simple directed graph with n vertices is a k-plex if the degree of each vertex of the graph is at least n − k. When k = 1, a 1-plex is a clique. Maximum k-plex problem aims to find a maximum vertex subset S of a given graph such that the subgraph G[S] induced by S is a k-plex. The applications and research on k-plex receive growing attention such as using k-plex to analyze social networks of terrorists, clustering and partitioning of graph-based data using k-plex], etc. The complement graph of a k-plex is a graph of maximum degree at most k- 1. Finding a maximum k-plex in a graph G is equivalent to find a maximum induced subgraph of degree bounded by k − 1 in the complement graph of G [73].

Overlapping Community Detection Algorithms

Social network is represented as a network graph. Detecting the communities involves finding the densely connected nodes. Overlapping communities are likely if a node is a member of more than one community. The clique percolation technique is a popular approach for analyzing the overlapping community structure of networks. The clique percolation method constructs up the communities from k-cliques, which correspond to complete sub-graphs of k nodes. Two k-cliques are considered adjacent if they share $k-1$ nodes. A community is defined as the maximal union of k-cliques that can be attained from each other through a series of adjacent k-cliques. Such communities can be best interpreted with the help of a k-clique templates where a template can be placed onto any k-clique in the graph, rolled to an adjacent k-clique by relocating one of its nodes and keeping its other $k-1$ nodes fixed. Thus, the k-clique communities of a network are all those sub-graphs that can be fully explored by rolling a k-clique template in them.

Clique percolation methods are categorized broadly into two types (i) directed clique percolation method (ii) weighted clique percolation method. The directed clique percolation technique defines directed network communities as the percolation clusters of directed *k*-cliques. A directed *k*-clique is a complete subgraph with *k* nodes on a network with directed links. The *k* nodes are ordered such that between an arbitrary pair of nodes, there exists a directed link pointing from the node with the higher rank towards the node with the lower rank.

The weighted clique percolation technique defines weighted network communities as the percolation clusters of weighted *k*-cliques. The geometric mean of link weights within a subgraph is called the strength of that subgraph. On a network with weighted links, a weighted *k*-clique is a absolute subgraph with *k* nodes such that the geometric mean of the $k(k-1)/2$ link weights within the *k*-clique is greater than a preferred threshold value, I [74].

The detailed description of the above-mentioned node-centric community detection algorithms are presented in the respective chapters.

SUMMARY

The main process of community detection relies primarily on the interactions users tend to have with other users on the network, along with the individual behaviors, forming some sort of communities characterized with dense connections within the community. In this chapter, basic definitions and terminology used in this thesis are presented. Also, concepts related to community detection and major four different categories of community detection approaches used in social network analysis have been introduced in this chapter. Various algorithms used in implementing community detection with experimental results are elucidated in the following chapters.