Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Babylon
College of Information Technology
Department of Information Networks

An Improved Leiden Algorithm Based on Cliques

A Thesis
Submitted to the Council of the College of Information Technology,
University of Babylon in Partial Fulfillment of the Requirements for the
Degree of master's in information technology- Information Networks

By

Baneen Ali Kadhim Mohammad

SUPERVISED BY

Prof. Dr. Ghaidaa A. Al-Sultany

2023 A.D. **1445 A.H.**

بسم الله الرحمن الرحيم

﴿ وَلَقَدْ آتَيْنَا دَاوُودَ وَسُلَيْمَانَ عِلْمًا وَقَالَا الْحَمْدُ

لِلَّهِ الَّذِي فَضَّلَنَا عَلَى كَثِيرٍ مِنْ عِبَادِهِ الْمُؤْمِنِينَ﴾

(سورة النمل 15)

Certification of the Examination Committee

We hereby certify that we have studied the thesis entitled (**An Improved Leiden Algorithm Based on Cliques**) presented by the student (**Baneen Ali Kadhim Mohammad**) and examined her in its content and what is related to it, and that, in our opinion, it is adequate with (Very good) standing as a thesis for the degree of Master in Information Technology-Information Networks.

Signature:
Name: Dr. Saad Talib Hasson
Title: Prof.
Date: / / 2023
(Chairman)

Signature:
Name: Dr. Huda Naje Nawaf
Title: Prof.
Date: / /2023
(Member)

Signature:
Name: Dr. Abdul Jalil Muhammad Majeed
Title: Asst. Prof.
Date: / / 2023
(Member)

Signature:
Name: Dr. Ghaidaa A. Al-Sultany
Title: Prof.
Date: / / 2023
(Member and Supervisor)

Approved by the Dean of the College of Information Technology, University of Babylon.

Signature:
Name: Dr. Wesam S. Bhaya
Title: Prof.
Date: / / 2023
(Dean of Collage of Information Technology)

Supervisor's Certification

I certify that the thesis entitled "(Enhancing Community Detection Using Maximal and Maximum Cliques into Hierarchical Algorithms)" was prepared under my supervision at the department of Information Networks / College of Information Technology/ University of Babylon by (Baneen Ali Kadhim) as partial fulfilment of the requirements of the degree of master in information technology- Information Networks.

Signature:

Supervisor Name: **Prof. Dr. Ghaidaa A. Al-Sultany**

Date: / /2023

The Head of the Department Certification

In view of the available recommendations, I forward the thesis entitled "Enhancing Community Detection Using Maximal And Maximum Cliques into hierarchical algorithms" for debate by the examination committee.

Signature:

Asst. Prof. Al-Harith Abdul Karim

Head of Information Networks Department

Date: / /2023

Dedication

To the one who sows to harvest, but I lack his applause for the joy of my achievement at this moment, you will stay my support and encouragement, you are in my heart

my dear father (may God have mercy on you)

The angel who inspired me with tenderness...my dear mother

I wish to express my gratitude to my beloved husband Safaa for your tremendous support and help

To my family my brothers and sisters. Their unwavering belief in my abilities

And to my second family

With All My Respect and Love…

Baneen Ali

Acknowledgements

First and foremost, I express my gratitude to Allah, God, as without divine care, I could not have even contemplated undertaking this study. I would like to extend my heartfelt appreciation to my supervisor,Prof. Dr. Ghaidaa A. Al-Sultany, for her unwavering support, expertise, and patience throughout this research journey. I am deeply grateful for her guidance and encouragement.

I also want to convey special thanks to my friends who stood by me throughout the entire study period, from the beginning of my academic journey until the completion of this thesis. Their encouragement and camaraderie have been invaluable and instrumental in keeping me motivated.

I am truly humbled by the assistance and blessings I have received during this research, and I recognize that this achievement would not have been possible without the collective support of those around me. Thank you all for being an essential part of this academic endeavor.

Abstract

In recent years, the rapid expansion of social networks has generated vast and complex networks comprising diverse interactions among users. Community detection has emerged as a pivotal approach to understand the structural organization of these networks by partitioning them into clusters of closely connected individuals.

This thesis presents a model that enhances the renowned Leiden community detection method through the integration of Maximal and Maximum clique algorithms. By leveraging clique analysis, our approach identifies relevant information from the network, leading to the division of the larger graph into smaller, more coherent communities.

The proposed model significantly increases the modularity ratio of the Leiden algorithm, thereby enhancing the precision of community detection. Rigorous experiments conducted on Facebook and Twitter datasets demonstrate the effectiveness of our model, with Modularity values increasing from 0.79 to 0.81 and 0.76 to 0.79, respectively.

Our findings underline the potential of this integrated approach in advancing community detection in social networks, offering valuable insights for better understanding user relationships and interactions in these dynamic virtual environments.

Table of Contents

List of figures

Chapter one
GENERAL INTRODUCTION

Chapter One
General Introduction

1.1 Overview

Analyzing various kinds of networks, especially social networks, has attracted a lot of attention in recent years due to the structural properties distinguishing these networks from the random ones. One of these distinguishing properties is the community structure.

Nowadays, individuals utilize social networks not just to stay connected with friends, but also to share their opinions, ideas, and even personal experiences like diaries. This activity greatly contributes to the distribution of information within the network. Researchers are actively studying and analyzing how information spreads and how users influence these networks [1] .

Networks are called social networks when there is communication and interaction between nodes within this network, as these networks consist of a group of nodes and links that allow people to connect and share their opinions, emotions, and ideas[2].

Social networks can be categorized into two types: static and dynamic social networks. Unlike static networks where the network's structure is determined by long-lasting friendship connections, many social networks are shaped by transient interactions among individuals[3]. The addition and removal of edges in a network have a significant impact on its structure over time. Examples of dynamic social networks include Facebook, Twitter, LinkedIn, Tencent WeChat, and Sina Weibo.

It is common for users to form groups within these social networks, allowing them to share thoughts and events with like-minded individuals. Understanding these subgroups provides valuable information about the overall structure of these networks. This phenomenon can be likened to the saying, "Birds of a feather flock together," where individuals with similar interests, personalities, or characteristics tend to associate with each other. This concept is illustrated in Figure (1.1). In a network, each node represents an entity, and the edges depict the communication or interaction between these entities[4].

Most networks exhibit a pattern where there are sparse connections between global groups and dense connections within local groups. This means that nodes connecting different communities have fewer edges, while nodes within the same community have more edges, thus defining the community structure of the network[5].

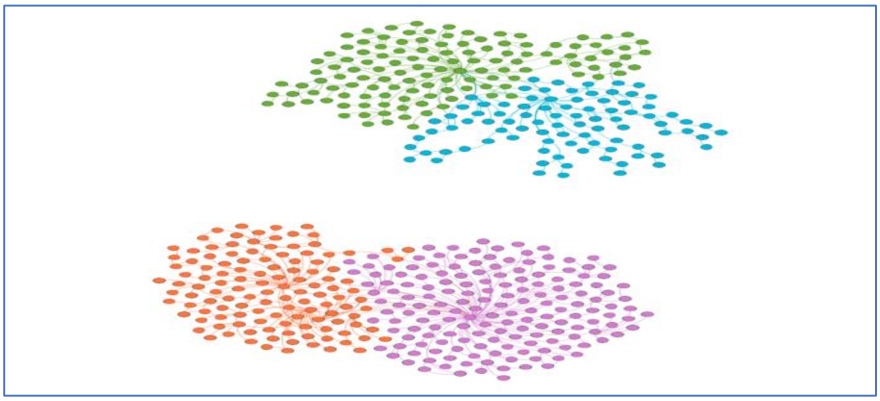


Figure (1.1) Example of communities in social network (68)

The process of communities detection in networks is considered one of the most important problems in the social network where it is considered Understanding the communities within a large network is vital for comprehending its structural and functional characteristics[6] . Community detection involves identifying clusters or communities within social networks. It is a crucial aspect of social networks and has various real-world applications. In essence, community detection involves dividing the network into clusters of closely interconnected nodes. Nodes within the same community should have strong connections with each other, while their connections to nodes outside the community should be weaker[7].

In addition, there are many environments and platforms for building communities where they can be built in the online social network where online social networks have found applications in various fields, such as marketing, healthcare , and political science, due to their impressive capability of reaching a wide audience quickly[8,9]. Complex networks, on the other hand, are a group of many interconnected nodes that communicate through various means. These nodes can represent individuals in real-life networks, such as social, biological, or technical networks. The interactions and connections between these individuals further define the nature of a networked system[10].

In a weighted network, the connections between nodes are assigned specific weights. In network science, a sparse network refers to one that has significantly fewer connections compared to the maximum possible number of connections within it. On the contrary, a dense network has a higher density of connections[11]. Community detection is a crucial problem in social networks, it like a graph theory where the goal is to partition a graph. Various researchers have attempted to identify optimal clusters of users within a network. There are multiple methods available to detect closely connected groups within the network's structure. Community detection serves as a useful tool to identify these groups as they can provide a condensed representation of the larger network, allowing for a comprehensive understanding of its structure[12].

1.2 Problem Statement

Community detection is a crucial challenge in social network analysis, particularly due to the rapid growth of these networks. Understanding the formation and evolution of links between users and the influence of community networks is vital. However, one of the significant problems faced in social networks is obtaining large communities that exhibit strong internal relationships and cohesiveness. To address this problem, the effectiveness of community detection can be enhanced by incorporating the principles of maximal and maximum cliques into the Leiden algorithm.

1.3 Aim of Thesis

The aim of this thesis is to enhance and develop a community detection strategy that focuses on achieving strong and cohesive communities within social networks. To accomplish this aim, the study proposes integrating maximal and maximum cliques into the Leiden algorithm to improve its performance in Identify communities with strong relationships.

By incorporating maximal and maximum cliques, which represent subgraphs with dense connections, into the Leiden algorithm, the thesis aims to enhance the accuracy and effectiveness of community detection. This integration is expected to result in the identification of more cohesive and tightly-knit communities within social networks.

The thesis seeks to contribute to the field of community detection by providing an improved strategy that can accurately identify and characterize cohesive communities. By achieving this aim, the study aims to advance our understanding of social network structures and dynamics, facilitating more insightful analysis and interpretation of social network data.

1.4 Research Objectives

The Objectives of this thesis can be divided into the following:

1. Clique Extraction: Initially, cliques were extracted from the dataset. This entailed identifying subsets of nodes that constitute fully interconnected subgraphs within the network. Subsequently, both maximal and maximum cliques were derived from the set of extracted cliques.

2. Leiden Algorithm Application: The Leiden algorithm was employed to ascertain communities within the network. Specifically, the obtained maximal and maximum cliques from the previous step were utilized as inputs to the Leiden algorithm. This approach aimed to optimize the detection of high-quality communities, while simultaneously reducing the computational effort and time required in comparison to directly applying the Leiden algorithm on the entire dataset.

3. Method Validation: The proposed methodology was rigorously assessed in the context of community identification. Through comparative analysis with alternative approaches, the effectiveness of the proposed method was evaluated. Encouraging results were obtained, demonstrating the ability of the enhanced approach to identify cohesive communities in a more reliable manner.

1.5 Related Work

Social networks have gained enormous popularity in recent years. Social networking sites allow users to connect with new people, discuss ideas with those who share their interests, and stay in touch with old friends and coworkers, but not only that, but they are also used in commerce to try to influence the opinions and desires of the users of these networks[13]. A person's appeal to other businesses or individuals seeking to advance an idea or sell a product increase with their level of influence.

This section reviews community detection algorithms[14].

the authors in[14] proposed a new method of the Leiden algorithm was developed to address the issue of poor community connectivity encountered in the Louvain algorithm. While the Louvain algorithm can result in node communities that become disconnected from each other, the Leiden algorithm utilizes an iterative process to ensure optimal assignment of community subsets and strong connectivity among community nodes. To improve both computational efficiency and the quality of detection, the algorithm incorporates the concepts of quick local movement and random movement of nodes to neighboring communities.

the authors in[15] proposed a method to decrease the temporal complexity of the Louvain algorithm. Their approach involves introducing a clique-Louvain algorithm as the initial step, specifically designed to identify cliques consisting of at least three members. The underlying idea is that every clique or node in the graph that is not part of a clique can be considered a community. Since the members of a clique are likely to belong to the same community, utilizing cliques as the focal point of communities enhances the algorithm's performance while maintaining a relatively high level of quality.

the authors in[16] It suggested that the enhanced Fast Louvain technique be used to increase the effectiveness of large-scale networks' detection. This study enhances iterative logic by switching from cyclic to dynamic iteration, which speeds up convergence and separates the local tree structure of the network. As a result, the algorithm is more effective. To speed up the computation, the network is divided up iteratively, the tree structure is added to the results of the partition, and the results are then optimized. The effect of community detection has been improved, and there is greater community aggregation. When it comes to partitioning and accelerating the process, the Fast Louvain method outperforms the conventional Louvain algorithm .

the authors in[17] proposed model was developed and exploited different techniques to enhance the Leiden community detection method. The first technique exploits hashtags to enrich tweets to improve the latent semantic analysis algorithm (LSA) performance. Second, Employing mention and retweet attributes to create networks used to detect communities. Third, a Support Vector Machine (SVM) is utilized for sentiment analysis to get users' opinions about a particular topic. Fourth, The topic modeling technique, Sentiment analysis, and retweet with mention were exploited to get a new weighted network. Finally, a new weighted graph can enhance the quality function of modularity for Leiden community detection method.

the authors in[18] presented a new method for identifying communities that takes into account user interactions and message content. This method represented interactions with users by retweet relations and messages contents with semantic similarity of users. They used topics exchanges between users to improve detects communities using traditional community detection methods .

in[19] a hierarchical clustering process was developed to identify communities by leveraging networks of user relationships and interests. The authors employed semantic analysis and the relationships of following/followers to calculate edge weights, thereby enhancing the community discovery process. The original network was transformed into a new network that consisted of undirected and weighted edges. The weights were generated using the direction and interest vectors from the original network, allowing for the determination of edge similarity. The hierarchical clustering algorithm was then applied to identify communities based on the edge-weighted similarity.

In[20] The authors of this study suggested a three-stage community detection methodology. The phases are community merging, label dissemination, and identification of the central node. Central node identification involves identifying central nodes based on node distances. Label propagation assigns the same color to nodes that exhibit maximum similarity. Finally, communities are merged if it results in an increase in modularity. The complexity of this approach is noted to be high for large networks.

In[21] the authors introduced Parallel Leiden, a parallelized version of the Leiden algorithm. By combining the active-nodes queue created by the Leiden algorithm with the PLM (Parallel Louvain Method) parallelization technique, we were able to implement local movement. The key concepts were to utilize compare-and-swap to check if nodes had already been added to the queue or not, and to employ a queue lock to prevent race circumstances while adding or removing nodes from the queue.

In[22] the authors presented an approach for the use of community discovery techniques utilizing statistical and visual analysis to aid in decision-making. To determine the most suitable community detection methods for a given dataset, Infomap and Louvain, two community detection algorithms, are tested on four real-world networks datasets with different features. The findings highlighted the parallels and discrepancies across algorithms that use statistical and visual analysis to aid decision-making.

In[23] the author proposed a method to identify nearly fully linked cliques in graphs by counting approximate maximal cliques. It was highlighted that starting with a triangle as a seed guaranteed the presence of at least one clique among its projected vertex sets. The objective was to employ heuristic search to locate all approximate maximal cliques that met the heuristic's constraints within each triangle's projection. The study utilized the A* search algorithm to select the best approximate maximal clique while eliminating less interesting nodes .

In[24] the Scalability problems were found when distributed and shared memory-based parallel algorithms were employed. A shared memory-based method using OpenMP produced a 4-fold speedup for many real-world networks. The speedup, however, is only constrained by the amount of physical cores that may be used in this arrangement. The ideal technique to utilize this system's physical cores is a distributed memory-based parallel method with a message passing interface. The code can be run on a sizable number of processors, according to the results. The biggest threat to building a scalable parallel Louvain method in a distributed memory environment is communication overhead.

The idea of the proposed model are combined two algorithms Maximal and Leiden algorithm to improve the discovery of the communities, where we first extracted the clique and then used the maximal clique algorithm to extract the maximal clique , then we entered the maximal clique into the Leiden algorithm and we noticed that the results were encouraging as we reduced the effort and time to discover these communities.

Table 1.1Summary of related work

No	Author(s)and year	Applied Algorithm	Dataset	Result
1	(Traag,Waltman and Van,2019) [14]	Leiden	live journal	0.76
2	(Elaf,2022) [15]	clique-Louvain algorithm	Facebook	0.60
3	(Zhang, Song and Feng,2021) [16]	Fast Louvain	Facebook	0.76
4	(Hyder ,2022) [17]	Leiden algorithm	Twitter	0.78
5	(Hoang , 2020) [18]	novel method	Twitter	0.75
6	(Li,Bai,Wenjun and Xihao ,2019) [19]	hierarchical clustering algorithm	Weibo	0.69
7	(You,Ma and Liu,2020) [20]	Infomap, Label propagation, fast greedy	Dolphin	0.52
8	(Fabian,2021) [21]	Parallel Leiden	Live journal	0.75
9	(Linhares ,Ponciano,Pereira,Rocha,Paiva and Travencolo,2020) [22]	Infomap and Louvain	High school in Marseille	0.79
10	(Abhishek,2021) [23]	A* search algorithm	Facebook	0.74
11	(Sattar and Arifuzzaman,2018) [24]	Parallelizing Louvain Algorithm	ego-Facebook	0.78
#	Our proposed system	Maximal and maximum with Leiden	Facebook Twitter	0.81 0.79

1.6 Thesis Outline

The remainder of this thesis chapters are organized as follows

- **Chapter 2:** Background and Literature Review Chapter 2 covers general information on social networks and provides an overview of community detection in social networks. It discusses the concepts and algorithms related to maximal clique and Leiden algorithm, setting the foundation for the proposed research.
- **Chapter 3:** Proposed Model for Community Detection Chapter 3 introduces the proposed model for community detection, which integrates the Leiden algorithm and maximum clique algorithm. It describes the methodology and techniques used to analyze the link relationships among users on Facebook. The chapter discusses the integration of these algorithms and presents the model in detail.
- **Chapter 4:** Experimental Results and Analysis Chapter 4 presents the experimental results obtained from applying the proposed community detection model to a Facebook social network. The chapter discusses the findings, analyzes the results, and provides insights into the performance and effectiveness of the model.
- **Chapter 5:** Conclusion and Future Work Chapter 5 summarizes the key findings of the dissertation and presents the conclusions derived from the research. It also provides recommendations for future work, suggesting areas of further exploration and improvement in community detection algorithms for social networks.

Overall, this thesis is structured to provide an introduction to the topic, a review of relevant literature, the development of a proposed community detection model, the analysis of experimental results, and concluding remarks with suggestions for future research.

CHAPTER TWO

Theoretical Part

Understanding Key Concepts and Models in Social Communities Detection

Chapter Two
Theoretical Framework

2.1 Overview

This chapter describes the Key Concepts and Models in Community Detection and Social Networks. so, at first of this chapter discusses the significant stages of the social network . Section 2.3 introduces the Structural Social Network Analysis, followed by Section 2.4, which focuses on Community detection. In section 2.5,the Clique based community detection is detailed. Finally, the evolution of communities is described in section 2.6.

2.2 Social Network

A real-world network can be visualized as a graph made up of nodes and links, where each link reflects how entities interact with one another and each node represents an entity or item. The network may be simple or intricate[25]. here in this research we will study social networks so the late 1990s is considered a date for the birth of social networks, as the social network Classmates.com appeared as the first social network to connect classmates with each other in 1995, and by 1997 SixDegrees.com came as a way to collect the bonds of relationships and direct links between users, It is noteworthy that these sites provided their users with a number of services. Such as sending messages, and despite the great popularity of these sites, they did not promise financial profit to their owners, and therefore they were destined to disappear[26].

During the period between 1999-2001, a group of social networks arose, and it was also destined to disappear, and with the advent of the year 2005, the American MySpace had occupied a prominent position among social networking sites, as it was indicated that the number of views of its pages had exceeded Google, and thus it is considered the largest network globally, and then Facebook came to compete with it in its global ranking[27].

Social networks are an abstraction of actual social systems, with individuals displayed as nodes and their interpersonal connections shown as links between nodes. A network's size and organization are determined by the number of its links and nodes, respectively.

Social networks are a group of nodes and links that can be denoted by graph G(V, E), where V is a set of nodes and E is a set of links that connect to create communities . The network is called a social network when there is interaction between nodes within this network[28].

A person's social network is shaped by their interactions and interpersonal connections with other members of society. Social networks serve as models and representations of interpersonal relationships. The rapid growth of the web has substantially increased the amount of user contact[29].

There are many social networking sites, like Facebook, Twitter, and others, are created to increase user interaction. As the number of contacts increases, it becomes more and more difficult to find these communications. People with similar preferences and tastes are more likely to become friends with each other.

Social media is a handy tool that makes it possible for people to expand their social lives in novel ways because it is challenging to make friends in real life but much simpler to find acquaintances with similar interests online.

Examining the patterns and properties of these real social networks can be done for several important purposes[30].

Social networks stand out for their ability to show a community structure. If a network's vertices can be separated into separate or overlapping sets of vertices, and if the number of edges inside each set is greater than the sum of the edges between any two sets, we say the network has a community structure. A hierarchical community structure is typically seen in networks having a community structure.

Due to the structural characteristics that set social networks apart from random ones, analyzing various types of networks has gained a lot of attention lately. The community structure is one of these distinctive qualities[31].

Finding strong associations or clusters inside a network is known as community detection. This is one of social networks' primary goals. It used graphs to analyze relationships between users on social networks to cluster users in different communities, which can be helpful in many applications in the real world[32].

Nowadays, researchers consider online social networks a necessary component of connections between people in the world. The network formation and spread of information have grown research rapidly that studies online social networks. The primary step for social network is the large-scale analysis of communication links, text contents, and information spreads. There are millions of active daily people on social networks. Online social networks are social structures of people or companies linked by relationships representing users' shared interests or ideas[33]. It consists of users defined by nodes that link and interact with other users or organizations by the edge . The structure of these networks is interesting because it provides insight into how people interact with one another. Friends tend to have a lot of mutual friends, which creates a social network. People with similar interests, opinions, and choices are more likely to connect in a social network, making different virtual communities or clusters. Communities are a set of more closely connected vertices than the rest of the network's vertices. The technique of grouping similar users into a cluster is known as community detection[34].

Social networks can be divided into static and dynamic social networks. Recently, a growing amount of attention has been paid recently to the identification of influential nodes in dynamic social networks[35]. Because the topology of static networks is determined by friendship links that slowly change over time, while the structure of many social networks is determined by transient interactions between entities, adding and removing of edges in a network significantly changes the structure of the network over time[36]. Each user in the network represents a node, and the edge represents the interaction from one user to another. These edges change over time as the interactions between users evolve. Social networks such as (Facebook, Twitter, LinkedIn, Tencent WeChat, and Sina Weibo) are networks that evolve dynamically over time. The process of selecting and attracting the most influential users is of utmost importance[37].

Dynamic networks are characterized by a variety of modifications that can co-evolve with the behavior of the nodes. In social networks, for example, Such changes take place as time passes and new members join the network, old members depart, and existing members establish or terminate relationships over time. These changes can lead to major shifts in the structure of the community, resulting in death, birth, growth, contraction, division, and merger of communities throughout the history of the network [42].

Many online social networks have emerged that allow users to communicate online, creating online user communities. These communities are used to exchange ideas, thoughts, collaborate, and make friends via social networks. Users can share and discuss a particular event or concept, such as a new movie in a cinema, a new product, a specific accident, or anything in the real world by posting text, images, and videos[38].Community detection is the procedure of identifying the clusters or communities on social networks. It is one of the most crucial components of a social network. It used graphs to analyze relationships between users on social networks to cluster users in different communities, which can be helpful in many applications in the real world[39].

2.3 Structural Social Network Analysis

In recent years, structural social networks have been emerged for visualizing and analyzing networks relationships. Researchers have found the importance of social network analysis in many areas such as biology, communication studies, information science, political science, social psychology, and computer science . Structural networks used graph theory to analyze relations in networks. Their main methods and techniques have been applied in large social network problems (i.e., Twitter, Facebook, Instagram, LinkedIn, Snapchat, YouTube)[33]. Social networks are considered complex networks because of the sparse and noisy network topology of social networks. In other words, node-to-node connections will not be random or completely regular.

Structural Social network Analysis (SSNA) refers to a method for extracting meaningful information from networks regarding their structures by measuring the relationship among network participants, including users, companies, URLs, or any other type of connected information processing structure [40].

In social network analysis, the composition of interactions between social entities is examined. Relationships formed can be intimate or formal, and they can range from passing acquaintance to tight ties. Links can show how information is shared, how people interact, and how things are similar. Graphs are frequently employed to depict the structure of these networks. A graph is comprised of two essential elements: vertices and edges. Graph theory describes social networks as nodes connected by edges, with nodes representing persons and edges representing their connections[40]. Depending on the application need, vertices can represent a wide range of individual objects such as(humans, governments, publications, companies, businesses, plants, and animals). On the other hand, an edge is a line that links two vertices and can represent a variety of interactions between distinct objects (e.g., friendship, communication, cooperation, acquaintances, and trade). Edges might be directed or undirected depending on whether the relationship is asymmetric or symmetric. A graph (network) is defined as a non-empty set V of vertices and edges, where E is the number of edges defined as G = (V, E). The following subsections introduce the most popular ways of finding communities in social networks[41].

2.4 Community Detection

The term "community" describes a social setting. People have a natural tendency to form groups, whether it's their workplace, their family, or their friends. A community is a collection of individuals who engage with one another more frequently than other users of the network, have similar interests, or consume material that is similar to one another[43].

In 1887 publication the book "Gemeinschaft und Gesellchaft" where the word "community" first appeared in this book. As of right now, there is no one definition of community that is generally used in social networks. Numerous definitions of community have been put out from various angles, however they can be broadly categorized into three categories: intuitive definition, functional definition, and definition from the algorithmic process[44]. and The earliest record of community detection study likely dates to 1927. Stuart Rice investigated voting patterns in minor legislative bodies during the time. He looked for blocs based on the degree of voting agreement between any two different groups, known as the Index of Likeness, and within any given group of individuals, known as the Index of Cohesion[45].

A community is a collection of items that are more similar to one another than the other entities in the dataset. Communities are created by individuals so that group members interact with one another more frequently than they do with outsiders. Measures of similarity or distance between the members of a group can be used to determine how closely related they are Similar to clusters in networks[46], social networks have communities. A node in a graph that represents a person may not only belong to one community or organization, but also to a large number of other connected or unrelated groups that are part of the network. Many researchers have identified and analyzed the community structure using methods from many branches of study. In order to evaluate the effectiveness of clustering in networks, the clustering coefficient, which is a measurement of how much a network's vertices tend to cluster together, is frequently utilized.

Two methods of community detection exist: overlapping and non-overlapping community detection. Non-overlapping or disjoint communities are those in which a node belongs to just one community and is unable to engage in any other communities. Additionally known as crisp assignment of nodes, overlapping communities refer to nodes that are associated with more than one community[47]. Modularity maximization, random walks, spectral clustering, statistical clustering, and differential equations are the five techniques we use to locate non-overlapping communities. Additionally, we have many techniques for identifying the overlapping community: clique percolation method, link partitioning and line graph, local optimization and local expansion, fuzzy detection, and agent-based and dynamic algorithm. Figure (2.1a) depicts a disjointed community structure split into two communities, indicated by the numbers {1, 2, 3, 4, 5, 6, 7, 8} and {9, 10, 11, 12,13} in two distinct hues. Each node in this instance is a member of a single community. Figure (2.1b) depicts an overlapping community structure that is divided into two communities, namely { 1, 2, 3, 4, 5, 6, 7, 8} and {8, 9, 10, 11,12,13} here, the node in green hue, {8}, it belongs to both communities.

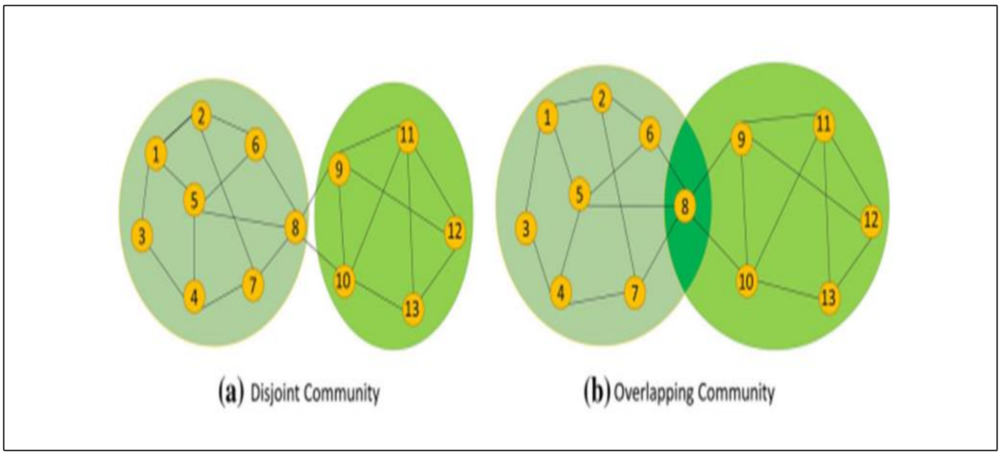


Figure (2.1) Categorizations of community detection algorithms (47)

Communities, or clusters, are typically subgraphs of nodes that are closely coupled to one another but sparsely related with others. The likelihood of nodes belonging to the same community increases with similar traits or behaviors, and vice versa. However, due to the complexity and diversity of networks, it is frequently very challenging to identify the precise organization or purpose of communities in many real networks. Therefore, techniques for community discovery would have a direct effect on our understanding of how networks are organized and operate.

We can identify modules with comparable structural or functional properties using community detection. Therefore, it is crucial to create algorithms that are precise and efficient for identifying communities. To identify community structures, a number of community detection techniques based on global or local information have been proposed[48].

The community structure of social networks is one of the most important characteristics of these networks because it can give us important information about the behavior of these networks. As a result, according to the research in [49], it can be characterized as follows: When the connections between nodes belonging to the same community are dense, the network G has a community structure, whereas nodes belonging to different communities are less connected.

Community detection on social networks is a serious issue. A graph theory that divides a graph is appealing . Finding the ideal user clusters has been the focus of numerous researchers. The practice of identifying closely knit groupings within a network's structure is known as community detection. It may be a helpful tool to find community because the communities can be a compressed version of the large network to obtain a complete view of the network[50].

Communities can be created in online social networks where Many people in online social networks such as Twitter and Facebook have billions of interactions between them, making analyzing these communications more complicated. Twitter has emerged as the most popular platform for research among online social networks because most data on Twitter is public by default and can be quickly accessed via the Twitter API[38].

2.4.1 Algorithm of Community Detection

Many community discovery algorithm are created by considering the entire network. Here are some of these algorithm.

The earliest community detection algorithm developed by Girvan and Newman (GN) is known as the GN algorithm. The betweenness value is primarily calculated for each edge of the graph. The edges with the greatest betweenness are then eliminated. Up until there are no edges remaining, these stages are repeated. Then, with increasing complexity over time, the level of communities with higher modularity is defined[51].

In (Bedi and et al,2016) One of the early methods to divide a graph was the Kernighan line algorithm. In order to reduce the total cost of all edges cut, it divides the network nodes with costs on edges into subsets of specified sizes. However, a significant drawback of this approach is that the number of groups must be predetermined[46].

(Bai et al.2017) the authors introduce a novel algorithm called OCDDP (Overlapping Community Detection Algorithm based on Density Peaks) for detecting communities. The algorithm consists of several key steps. Firstly, a distance matrix computation approach is employed. Then, a three-step technique is utilized to select community cores and determine the cluster centers. Finally, a node allocation mechanism based on membership vectors is developed. The efficiency of OCDDP is demonstrated to be superior when dealing with simple networks compared to existing techniques. Additionally, OCDDP exhibits strong performance even when applied to complex networks[52].

In (Song and et al.

2016) used the Bat Algorithm (BA) to detect communities. The PSO (particle swarm optimization) technique and the simulated annealing algorithm are combined in BA, which has great convergence and global search capabilities. BA replicates the frequency, emission rates, and loudness of bats foraging based on their behavior when using echolocation. By varying the pulse frequency, bats may alter the wavelength . Bats can locate the target when the wavelength matches the size of insects. BA uses a random flight operation to provide a random solution in order to avoid the PSO flaw of slipping into the local optimum. Because BA's local search operation resembles that of the simulated annealing process, it converges more rapidly.

BA is more exceptional since it combines the benefits of the simulated annealing process and the PSO algorithm[53].

(Zhou et al. 2017) the authors introduced the AR-Cluster algorithm, which aims to detect communities in social networks. This algorithm utilizes a unique collaborative similarity metric to cluster nodes based on calculated similarities among vertices. The clustering process is implemented using a K-Medoids framework. Experimental results indicate that the AR-Cluster approach outperforms three other techniques (W-Cluster, SA-Cluster, IGC-CSM) in terms of performance. However, it should be noted that the AR-Cluster approach may face challenges when applied to large social networks[54].

in(Blondel et al.2018) introduced The Louvain method for community detection the Louvain method greedy update adjusts node by node and uses the best neighboring community to optimize the modularity function gain. The algorithm then aggregates the result partition and repeats the process until no new communities are developed. The Louvain technique is fast and efficient, but it still gets stuck at local optima and might cause communities to become separated[57]. The Leiden method resolves the Louvain problem by adding a refinement step, but it still depends on greedy local updates and is vulnerable to local optima[55].

2.4.2 Leiden Algorithm

Detecting the community is significant work when investigating complex social networks. Leiden was built based on the Louvain algorithm that overcomes the problem of community connection badly. The Louvain method may cause to discover communities that are arbitrarily badly connected. It can detect internally disconnected communities where one part of an internally divided community can only connect with another through a link outside the group (see figure 2.2). Leiden is an iterative strategy that ensures all subsets of communities are properly assigned, and nodes within communities are firmly connected[56].

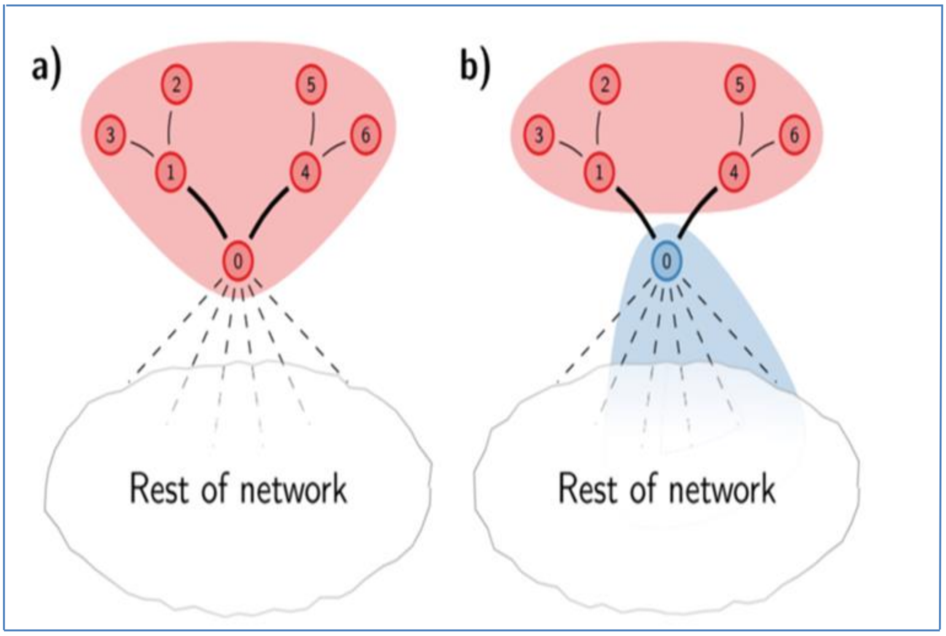


Figure (2.2) Louvain arbitrarily poorly connected (56)

The Louvain algorithm is modified by the smart local move algorithm, which forms the basis of the Leiden technique. The Leiden algorithm also uses ideas of speeding up local node movement and moving nodes to random neighbors. These modifications to the Louvain algorithm are the most promising ones. The Leiden method is straightforward to comprehend and straightforward to use. Compared to the Louvain algorithm, it has a lower time complexity. It consists of three stages: the local moving of nodes, refinement of the partition, and network aggregation based on the refined split.

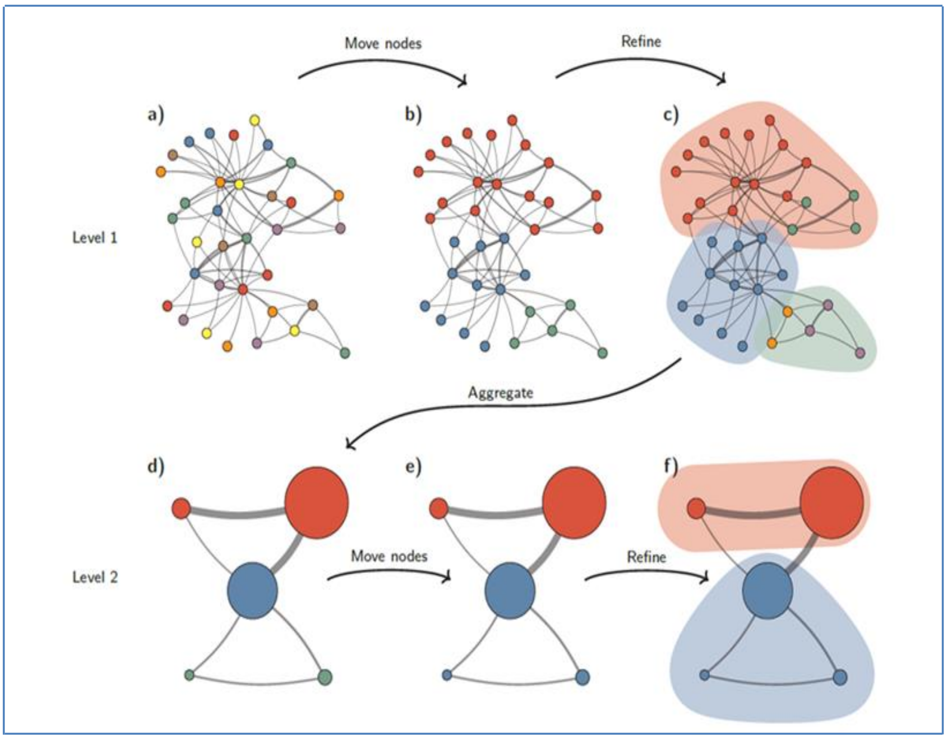


Figure (2.3) Leiden Method (56)

The Leiden algorithm's initial stage starts with a community of singleton nodes. The nodes then move across communities in seek of the optimal partition. The algorithm then performs a refined stage that aggregates the network. A quick local movement algorithm was used by the Leiden algorithm, and it only visited nodes whose vicinity had changed. The Louvain approach, on the other hand, repeatedly examines each node in a network. The algorithm is visualized in figure(2.3). The refinement stage might split a community into multiple communities when they have poorly become connected after a node move, increasing the connectedness of the remaining communities[57]. Nodes are not always greedily merged with the community that provides the most significant increase in the quality function. Instead, a node may be combined with any community that improves the quality function. The community that a node is merged with is selected at random. The higher the quality function increases, the more likely a community will be chosen.

Recently, the Leiden algorithm has been used widely in many application biological areas [58] because of its guaranteed connection, hierarchical partitioning, and high modularity. The method maximizes each community's modularity, which measures the quality of distributing nodes to communities by comparing the density of edges with a set of nodes to how they would be connected in the network.

2.5 Clique Based Community Detection

Because numerous definitions might be offered for various purposes, it is actually tough and challenging to propose a thorough and exact description for a clique in social networks. To examine experimental data, Luce and Perry established the clique model in 1949. A clique was described as a subset of vertices in a graph that contained all possible edges. Specifically, it refers to a subgraph in which any two vertices are close to one another. Due to its strong relationship to the fundamental idea of cohesive subgroup, the introduction of the clique model garnered a lot of interest in many disciplines[59] .

A clique is a complete subgraph of a larger graph, according to one definition. Cliques are collections of vertices that are close to one another, presuming that there is a distance function between any two vertices. The principal application of this definition is used in data clustering and the clique is a topological and structural characteristic of the graphs.

clique finding techniques have been presented For weighted and unweighted, directed and undirected, [60]and the graphs were obtained from combining them, Cliques can be categorized into two groups based on how much they overlap with one another and how much they don't. The majority of the graphs linked with social networks it overlap with one another In other words, a vertex may be a part of many cliques. Examples of overlapping cliques in graphs include the Facebook graph and graphs produced by models.

Cliques are a particular kind of group that allow for full communication between all of its members; a clique is maximal if it is not a part of a larger clique, and it is known as the maximum if there is no other clique in the network that is larger than it. The symbol w(G) indicates the size of the greatest clique in graph G, also known as the clique number. are a clique's structural attributes are Distance, diameter, domination, degree, density, and connection. A clique's structural characteristics can be used to define it.

The clique is a subset of undirected graph vertices where each pair of different vertices is neighboring. The clique is made up of paired neighboring nodes that are connected by edges, each of which represents a full sub-graph[61].

Clique is used to identify shared nodes in a graph to create a community. Finding the communities in a network depends heavily on cliques. The fact that the nodes in the cliques are totally connected to one another which indicates that they must belong to the same community[62].

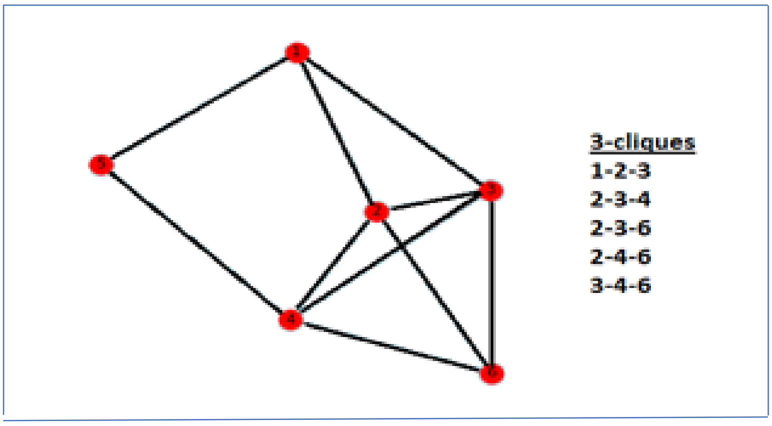


Figure (2.4)A Graph with 5 cliques of size 3

2.5.1 Maximum Clique

Finding the biggest clique in a given graph is the goal of the maximum clique problem. A lot of research attention has been paid to this well-known NP-hard problem. The clique number, denoted as ω(G), represents the number of vertices in a maximum clique of graph G. Additionally, the concept of maximum clique transversal refers to a subset of vertices that must be present in every maximum clique of the graph[62].

In Fig (2. 5) nodes {1, 2, 4, 5}a maximum clique of size four means that No clique larger than four nodes can be generated.

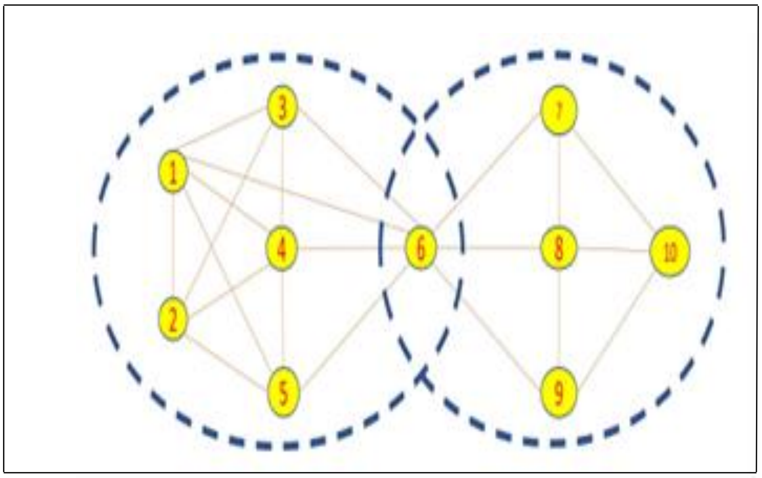


Figure (2.5)Community structure (61)

2.5.2 Maximal Clique

Graph analysis' essential purpose of community detection is to find collections of closely related vertex groups. One popular approach for community detection is the utilization of maximal cliques, which are subsets of vertices where every pair of vertices is directly connected and also known as a clique that can be extended by including one more adjacent vertex , is a clique that can be enlarged by adding one more adjacent vertex and is a component where each node is connected to every other node in the sub-graph, i.e., the degree of connectivity is 100%. Maximal cliques have various applications in data analysis, community detection, bioinformatics, and data mining.[65] .

The densest substructure in the graph is called the maximal clique and is perhaps the most widely employed as a potent community discovery method.

A graph's maximal clique is a portion of the graph in which the clique's characteristic is not lost when a node is added. An independent vertex set that can be enlarged to another independent vertex set by the addition of any vertex in the graph is said to have a maximal independent vertex set and is a graph where maximal cliques are the nodes, and an edge connecting two maximal cliques if they share a vertex.

2.6 Community Evaluation

The community structure of social networks is one of the most important characteristics of these networks because it can give us important information about the behavior of these networks. As a result, it can be characterized as follows: when the connections between nodes in the same community are dense, the network G has a community structure, whereas nodes belonging to different communities are less connected. Modularity Q is a metric that is frequently used to describe the robustness of a network community structure [66].

Modularity is often used in optimization approaches for discovering community structure in networks.. After executing the Leiden algorithm, modularity was obtained as in equation (2.1) [67}]

$$Q= \frac{1}{2M}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2M})\delta(c_i c_j) \tag{2.1}$$

Where

- A_{ij} denotes real no. of edges in community
- k_i and k_j denotes the aggregate of degrees of nodes in each community
- M indicates the total number of edges in the dataset
- C_i is the community of node i
- δ = 1, if same community, otherwise 0

Furthermore, in a realistic social network, modularity Q is typically between 0.3 and 0.7. The social network's community structure grows stronger as modularity value increases.

Chapter three

Enhancing Community Detection in Social Networks

Chapter Three
Enhancing Community Detection in Social Networks

3.1 Overview

This chapter presents a comprehensive exploration of techniques aimed at enhancing the community detection method in social networks. To achieve this goal, a new method is proposed, combining the strengths of the methods discussed in the preceding chapter. The integration of the "maximal clique" algorithm and the "Leiden algorithm" offers a promising strategy to improve the accuracy and efficiency of community detection in large and complex social networks. The chapter discusses the significant stages of the proposed model, outlining the step-by-step process of merging these algorithms to effectively identify communities.

3.2 Proposed Model

The main objective of the proposed model is to improve community detection in large and complex networks, with a focus on reducing the computational time and effort required for this task. To achieve this goal, the model incorporates two key algorithms, namely the "maximal clique" algorithm and the "Leiden algorithm." The overall framework of the model, as depicted in Figure (3.1), consists of several crucial steps designed to efficiently detect communities in the network.

1.Preprocessing Stage: The first step in the proposed model is the preprocessing stage, where the raw data is ingested by reading the data file.

Ensuring data integrity is critical for obtaining accurate results; thus, the model checks for the presence of missing values and applies necessary data cleaning techniques to prepare the network data for subsequent analysis. By conducting thorough data preprocessing, the model aims to enhance the quality of the input data, which directly impacts the community detection process.

2.Finding the Maximal and Maximum Clique: Following the preprocessing stage, the model proceeds to apply the "maximal clique" algorithm to the preprocessed network data. The maximal clique algorithm is utilized to identify and extract all the maximal cliques present in the network. A maximal clique is defined as a complete subgraph where each node is directly connected to all other nodes within the subgraph, and no additional node can be added without violating the completeness property. Identifying these maximal cliques helps to reveal densely connected regions in the network. Additionally, the model identifies the "maximum clique" from the list of all extracted maximal cliques. The maximum clique corresponds to the largest clique in terms of the number of nodes it contains. This step is valuable as it provides insight into the most densely connected region of the network, potentially indicating a core community.

3.Finding Communities by Combining Algorithms: The third and crucial step involves community detection by combining the results obtained from both the maximal clique algorithm and the Leiden algorithm. The Leiden algorithm, known for its effectiveness in identifying communities in large-scale networks, is employed to further refine the community detection process. The combination of algorithms allows for a comprehensive approach to community detection, leveraging the strengths of both methods. There are various ways to achieve this combination, such as using the maximal clique as a seed for the Leiden algorithm or merging the communities detected by both algorithms to obtain a more refined and accurate community structure.

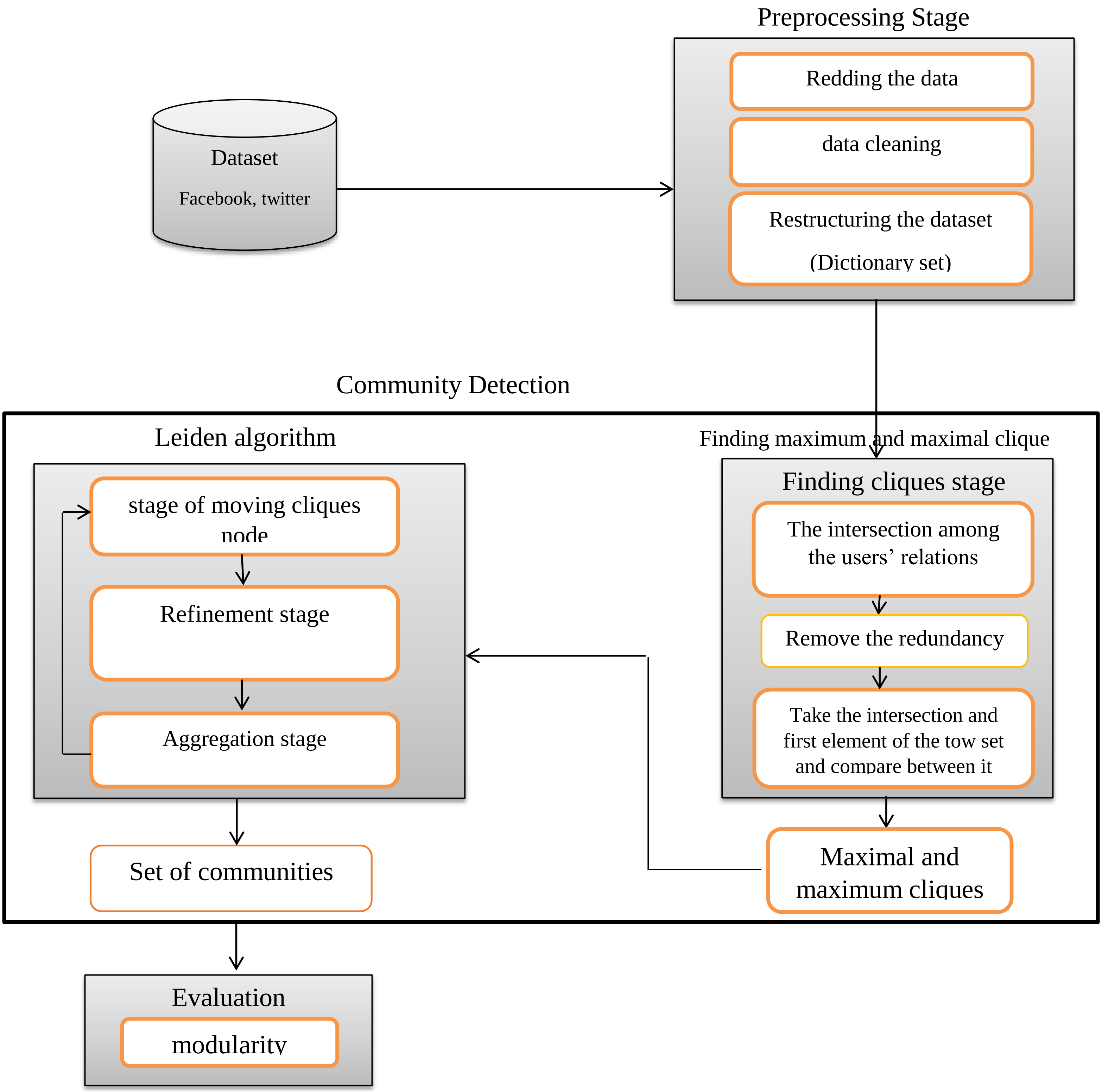


Figure (3.1)The proposed model

3.3 Dataset Description

A crucial aspect of evaluating the proposed model for enhancing community detection in social networks is the use of appropriate and well-known datasets. These datasets contain structural information of social networks, with individuals represented as nodes and their connections as edges. The two real-world information spreading datasets utilized to evaluate the proposed model are Facebook 2005 and Twitter 2010.

3.3.1 Facebook Dataset

The Facebook dataset captures the social network of Facebook friends and the posts made by each user during the period from September 25, 2006, to January 22, 2009. The dataset is represented as a text file, specifically a Friendship links file. Each Facebook user is identified by a unique User_ID, and the friendships between users are denoted as undirected links. The dataset contains 46,952 nodes and 876,993 edges, signifying the social relations between users. The contents of the dataset files are described in Table 3.1.

Table 3.1: Fields of raw Facebook relation

Fields	**Definition**
User_ID	Unique identifier of a Facebook user
Friend_ID	Social relation between User_ID & Friend_ID

3.3.2 Twitter Dataset

Twitter, a microblogging platform, allows users to share 140-character tweets, which can be liked, commented on, retweeted, or mentioned by other users. Users can follow each other, creating one-way links. The Twitter dataset used in the evaluation comprises approximately 855,650 nodes and 1,565,117 edges. It consists of two fields: Tweet_ID, representing the user's ID, and Retweet ID, indicating the relationships that connect users with their friends.

Table 3.2: Fields of raw Twitter relation

Fields	**Definition**
Tweet_ID	Unique identifier of a Twitter user
Friend_ID	The ID of the user and the friend's relation

Twitter networks are represented as interactive graphs, where nodes represent users, and edges denote relationships between users. For instance, nodes can represent individuals or blogs, while edges can represent friendships, hyperlink relationships between blogs, or any other relevant connections. The social network is represented as G (V, E), where V denotes the set of users (nodes), and E represents the set of connections between users (edges). The construction of the social dataset is based on the nature of relationships among users, as illustrated in Figure 3.2.

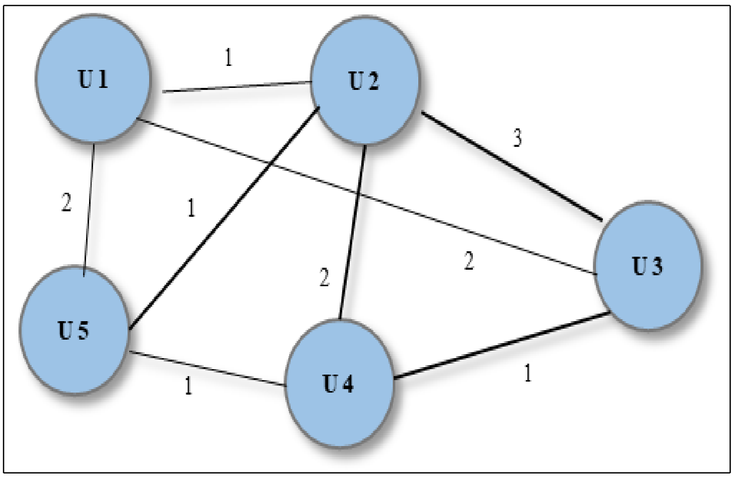


Figure (3.2)Sample Graph Network

3.4 Preprocessing Stage

The preprocessing stage plays a crucial role in transforming the raw data from TXT files into a structured and readable format required for the subsequent steps of the proposed model. The words, sentences, and networks detected during this stage serve as the foundational components for all further processing, making it an integral part of the entire model. By converting unstructured data into a more suitable format for analysis, a satisfactory preprocessing process significantly enhances the accuracy of the social network analysis results.

In the case of the Facebook dataset used in this research, the preprocessing process begins by reading the data file. The data is then thoroughly cleaned, ensuring that there are no missing values or any other irregularities that require processing. The objective is to obtain a clean and well-organized dataset format that can be readily understood and utilized by the community detection algorithm.

The main tasks performed during the preprocessing stage are as follows:

1. **Read dataset**
   - In the Facebook dataset, each user is uniquely identified by an identifier associated with them in the data files ,and the data must be converted into dictionary set before used it.
   - In the Twitter dataset, users are identified using unique ID numbers.
2. **Relationship Extraction**
   - For the Facebook dataset, the preprocessing stage involves extracting the source and target nodes from the relation information text file. This step is crucial in simulating the diffusion process within the network.
   - In the Twitter dataset, the preprocessing process focuses on extracting the usernames of users who tweeted and those who retweeted. This information is essential for analyzing the spread of information in the network.

By completing these tasks in the preprocessing stage, the data is transformed into a structured and standardized format suitable for subsequent analysis. The cleaned and organized dataset enables the community detection algorithm to accurately identify and analyze the social network's communities. Ultimately, a robust preprocessing stage contributes significantly to the overall quality and reliability of the social network analysis results.

3.5 Enhancing Community Detection with Clique And Leiden Algorithm

The community structure of social networks is a pivotal characteristic that provides crucial insights into the underlying organization and relationships among nodes. Dense connections between nodes within the same community signify the presence of a strong community structure in the network, while nodes belonging to different communities exhibit fewer connections, highlighting their distinctness. Community detection represents one of the most significant challenges in social network analysis, with extensive research aimed at identifying large communities with robust internal relationships. Despite these endeavors, achieving satisfactory results has remained elusive. Therefore, in this research, we propose an approach that combines the "Maximal Clique" algorithm and the "Leiden algorithm" to enhance the discovery of communities, while also reducing the effort and time required for community detection in social networks. The integration of these algorithms aims to capitalize on their individual advantages and overcome their respective limitations, ultimately leading to more accurate and efficient community detection results.

3.5.1 Finding Cliques

Cliques play a significant role in identifying communities within a network. A clique is a group of nodes in a graph that are completely connected to each other, indicating that they likely belong to the same community. In the context of large networks, such as social networks, finding cliques is crucial for community detection. The existence of completely connected subgroups helps in efficiently identifying communities and their relationships in the network. Cliques offer several benefits, including time-saving, increased productivity, reduced network overload, accelerated analysis and decision-making, and insights into individual behaviors and interests. They also help in managing relationships within social networks.

In this research, both the Facebook Dataset and Twitter Dataset are utilized. The Facebook data consists of 46,952 nodes and 876,993 edges, representing the friendship relationships among users on Facebook.

The basic steps for correctly extracting cliques are as follows:

1- Firstly, a function is created to compare the groups extracted in the form of a directory. This function extracts the intersection between these groups, ensuring that shared nodes are identified accurately.

2- After extracting the intersection between the groups, the next step is to eliminate any repeated elements resulting from the previous function. This ensures that the dataset is free of any duplicate entries.

3- Finally, the correct cliques are extracted from this cleaned dataset, where each node is in communication with every other node in the clique as explain in algorithm(3.1).

Algorithm 3.1 Finding Clique

```
input new Network G=(V,E)    //G data set ,V is the source of dataset and E is the target
output :the number of cliques in network G
Begin:
1.      D       //red the dataset G and convert G  into dictionary node (k, u)
2.          for k, u in D         //where the k is the name of set in D and the u is the element of set
3.                  if k == k1: continue
4.                  C= list(set(u) ∩ set(u1))
5.            Then
6.                R= set (k,u)
7.      function check the R is validate  clique
8.         for i in R
9.                  If i in G
10.         x= validate clique
11.             else
12.          x= not validate clique
13.           end function
            End
```

3.5.2 Maximum And Maximal Clique Algorithm

In graph theory, maximum and maximal cliques hold vital importance in various applications, including community detection, network analysis, and optimization problems. The clique number of a graph G, denoted by ω(G), is the number of vertices in its maximum clique. In other words, a clique is considered maximum if its size is the largest among all cliques in the graph.

After the process of extracting cliques as described in the previous section, the research proceeds to extract the maximum clique. This is achieved by calculating the elements of each clique separately and identifying the clique that contains the most number of elements, making it the maximum clique in the network as explain in algorithm (3.2).

Subsequently, the research focuses on extracting the maximal clique based on the previously extracted cliques. The algorithm identifies the cliques in which the addition of a node does not violate the clique's properties, and thus, the clique remains maximal as explain in algorithm (3.2).

Algorithm 3.2 find maximal and maximum Clique

input the cliques in algorithm 3.1

output :Maximal and maximum cliques

Begin:

1. R //the output of set of cliques finding in algorithm 3.1
2. Function extract the maximal from the cliques R
3. for i in R // where i is the set 1 in R
4. for j in R //where j is the set 2 in R
5. if i == j: continue

```
6.            c = i ∩  j
7.               F = (Print i and j )are maximal cliques
8.   End function
9.   Function extract the maximum from the cliques R
10. For (i , j) in R
11. if i > j          //i is the clique1 and j is clique2 in R
12. N= i              //the maximum clique
13. else
14. N=j
15. end function
     End
```

3.5.3 Utilizing Leiden Algorithm

By combining the power of both maximum and maximal clique algorithms and incorporating them into the Leiden algorithm, our research aims to significantly improve community detection in social networks. These integrated approaches leverage the strengths of cliques, maximizing and maximizing, to unveil densely connected subgroups and enhance the overall community detection process.

The Leiden algorithm, known for its effectiveness in community detection, becomes even more robust with the inclusion of cliques. After the extraction of cliques as described earlier, we proceed to incorporate them into the Leiden algorithm as part of its community assignment process.

The "Leiden algorithm" is recognized as a powerful tool for community detection, known for its efficiency and effectiveness. It focuses on partitioning the network into distinct communities where all subgroups are assigned efficiently. Unlike the Louvain algorithm, the selection of nodes in the Leiden algorithm is not random. Instead, nodes are placed in a queue, and modularity is calculated for each node to determine its optimal community assignment.

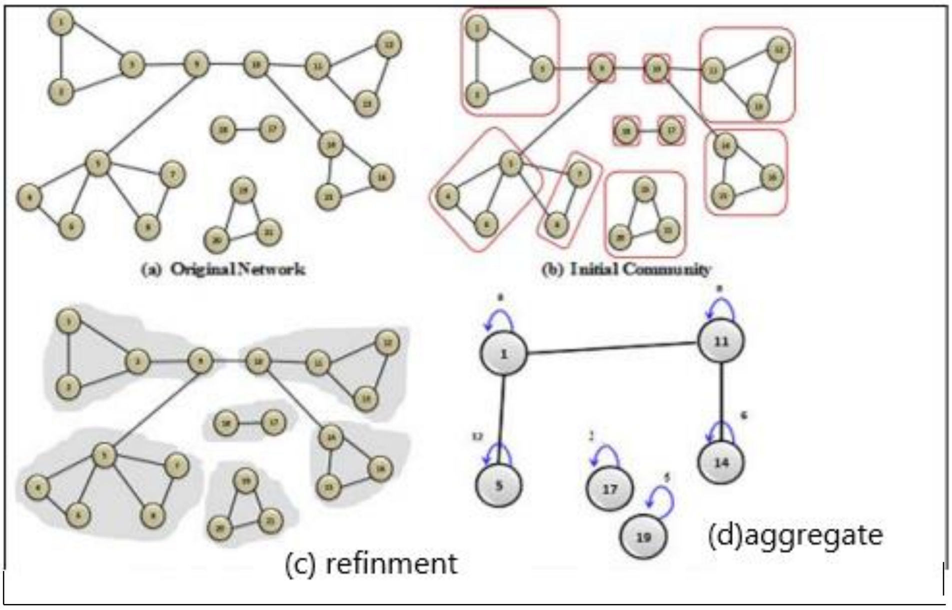
The Leiden algorithm involves three main stages:

1.Local Moving: Modularity is calculated for each clique (both maximal and maximum cliques), treating each clique as one node. Nodes are moved to different communities if there is an increase in modularity. Only adjacent nodes that do not already exist in the community are considered and placed at the end of the queue.

2.Refinement: This stage may split a community into multiple communities to improve the connectedness of the remaining communities. Nodes are not necessarily greedily merged with a community; instead, any community that enhances the quality function may be paired with a node.

3.Aggregation: Nodes within the same community are combined into one large node.

Generally, the key point of our proposed algorithm is to identify communities in networks by leveraging the inherent cliques. Given the graph G = (V, E), where V represents a set of vertices, and E represents the edges between vertices. Figure 3.3 (a) illustrates a simple network with a coherent structure of different sizes as an example. The algorithm proposes the initial step of finding maximal cliques involving at least three members (see Figure 3.3(b)). Each clique or a node (if not a member of any clique) in the graph is considered a community. It is likely that the members of a single clique practically belong to one community. The first step of the algorithm involves enter some of cliques to the queue calculating the modularity gain. Next, the algorithm moves the vertices or cliques accordingly. These two steps are repeated until the modularity value no longer improves, thereby generating hierarchical communities.



Figure(3.3) maximal with Leiden Algorithm steps

In practice, the algorithm satisfies the stop criteria when there is no further improvement in the value of the modularity. Algorithm (3.3) outlines the steps in detail.

Initially, the algorithm considers the existing communities, whether they are cliques or individual nodes. One community is chosen based on a queue, and the modularity gain of merging it with a neighboring community (clique or node) is computed.

Subsequently, the community chosen is moved into the neighbor that achieves the maximum modularity gain, as illustrated in Figure 3.3(c). This step ensures that the community structure is optimized for higher modularity.

Finally, the traditional algorithm's original process resumes, where each resultant community is represented as one node with a self-loop. The self-loop indicates internal connections among nodes within the same community, while edges represent external connections between nodes from different communities, as shown in Figure 3.3(d).

Throughout this process, the algorithm iteratively works to improve the modularity of the community structure. Once no further improvement in modularity is possible, the algorithm stops, and the resulting community structure is considered as the final outcome.

Algorithm 3.4 Clique based Enhanced Community Detection

Input: the maximal and maximum cliques in algorithm 3.2

Output: Communities Partition

Begin:

```
1.     R     //the output of set of cliques finding in algorithm 3.1
2.     G     //the output of maximal and maximum cliques in algorithm 3.2
3. maximal= find_cliques(R)
4.   maximum =  find_cliques(R)
5.     Do
6.        L ← Fast _Local _Move(G, L)
7.        L == N (G)     //Every community is one Node
8.         If not done then
9.            pref  ←Refine_Partition (G, L)
10.                G ←Aggregate_cliques_refined (G,  pref)
11.                L ← Partition(V(G))
12.        End if
13.    While Not Done
14.    Return Communities L
  End
```

3.6 Advantages Of The Proposed Approach

By combining the "Maximal Clique" algorithm and the "Leiden algorithm," our proposed approach offers several advantages:

1. Improved Accuracy: The combination of algorithms enhances the accuracy of community detection by effectively identifying cohesive communities with strong internal relationships.

2. Enhanced Efficiency: The use of cliques as initial building blocks reduces the search space, making the community detection process more efficient, especially for large and complex networks.
3. Reduced Time and Effort: The integrated approach streamlines the community detection process, reducing the time and effort required to discover meaningful communities within social networks.
4. Deeper Insights: The proposed approach provides deeper insights into the community structure, interactions, and behavior of individuals within the social network.

CHAPTER FOUR

Experimental Results And Discussion

Chapter Four
Experimental Results And Discussion

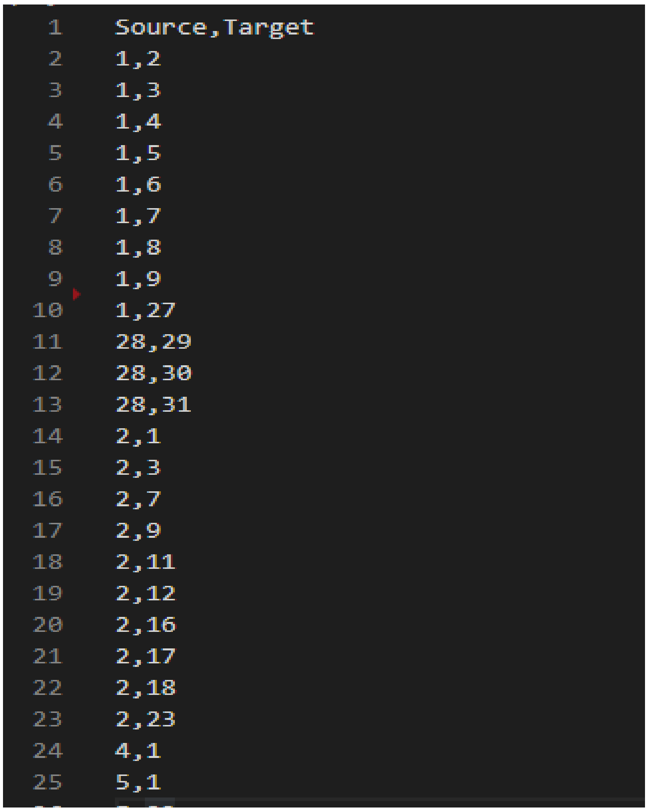
4.1 Overview

This chapter presents the implementation and experimental results of our proposed model, which has been tested on both the Facebook and Twitter datasets. The process includes preprocessing the datasets, finding cliques, calculating maximal and maximum cliques, and using the Leiden algorithm for community detection. We will investigate the Leiden community detection results and compare them with other methods on the datasets.

4.2 Data Preprocessing

As mentioned in Chapter 3, we used the Facebook dataset to build and test the proposed model. The Twitter dataset was also used to validate the model. Since the two datasets share similar features, we will review the features of the Facebook dataset in this section.

The Facebook dataset is one of the widely used social media networks worldwide. We utilized a clean dataset without any missing values, enabling us to use it directly without preprocessing. This dataset represents a social network of Facebook friends from September 25, 2006, to January 22, 2009. It consists of 46,952 nodes and 876,993 edges. For our research, we used 250 nodes and 2426 edges, as illustrated in Figure (4.1).



```
Source,Target
1,2
1,3
1,4
1,5
1,6
1,7
1,8
1,9
1,27
28,29
28,30
28,31
2,1
2,3
2,7
2,9
2,11
2,12
2,16
2,17
2,18
2,23
4,1
5,1
```

Figure (4.1) Facebook dataset

Twitter is another widely used social media network globally, where one-way links allow users to follow others without mutual exchange. This dataset represents a social network of Twitter friends from 2010. Similar to the Facebook dataset, we used 250 nodes and 699 edges from the Twitter dataset, as shown in Figure (4.2).

```
Source,Target
12,13
12,14
12,291
12,295
12,409
12,418
12,422
12,648
13,107
13,12
13,20
13,291
13,295
13,422
13,586
14,12
14,380
14,409
14,586
20,107
20,13
20,291
20,295
20,414
20,418
20,422
20,586
```

Figure (4.2) Twitter dataset

4.3 Finding Clique

We followed several steps to find cliques of the Facebook dataset:

1.We converted the dataset into sets or dictionary.

2. A function was created to compare the extracted groups in the form of a directory, extracting the intersection between the groups.

3. We eliminated the repeated elements resulting from the previous function.

4. Finally, we extracted the correct cliques from this dataset.

Using 250 nodes, and we obtained 105 overlap cliques, where the minimum clique size is 3 nodes some of which are shown in Figure (4.3).

```
These are the vaild cliques in the data :
[(1, 5, 22, 27),
 (1, 2, 9, 23),
 (1, 10, 13),
 (1, 14, 21),
 (1, 21, 25),
 (1, 26, 27),
 (28, 29, 31),
 (32, 33, 35, 37),
 (38, 39, 43, 44, 46, 47, 48),
 (38, 40, 42),
 (38, 41, 44, 46),
 (8, 50, 51),
 (7, 50, 53),
 (64, 68, 70),
 (64, 67, 68, 71, 72),
```

Figure (4.3) The extracted cliques for Facebook dataset

Here are the graph of some Facebook dataset cliques , where we took approximately 250 nodes and we have 105 cliques.

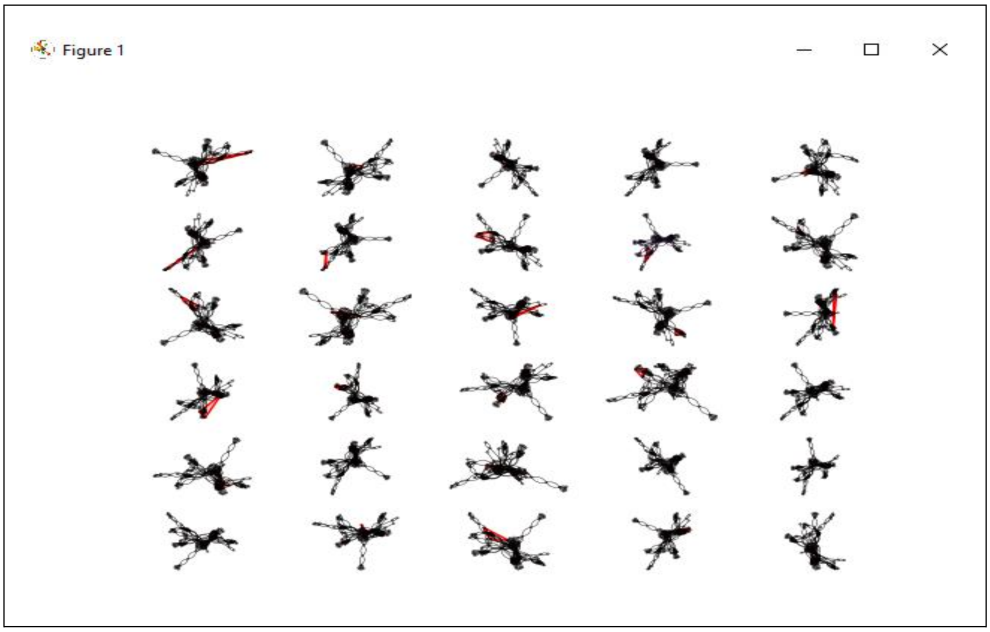


Figure (4.4) graph of cliques for Facebook dataset

Similarly, we used 250 nodes from the Twitter dataset, and we obtained 92 cliques.

Here The graph of some cliques of twitter dataset when we took approximately 250 nodes and we have only 92 overlap cliques.

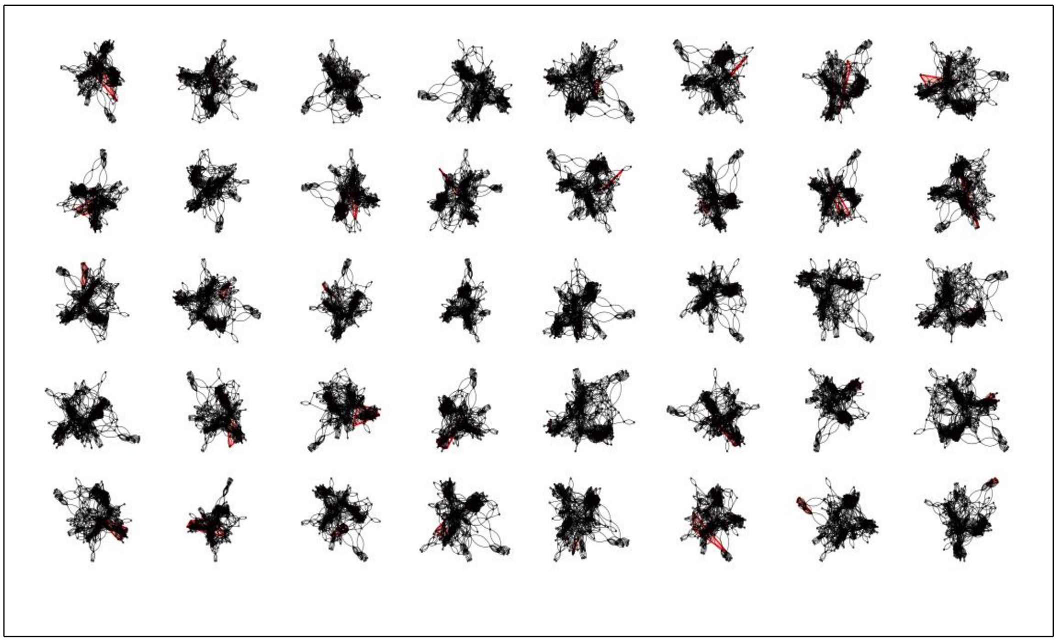


Figure (4.5) graph of cliques for Twitter dataset

4.4 Finding Maximum And Maximal Cliques

At this stage of work, after 250 nodes were used for Facebook and Twitter dataset, we calculated the cliques for these data, where we got 105 clique for Facebook Dataset and 92 clique for Twitter Dataset.

After this stage, we extracted the maximal clique from the previous cliques, where we got 102 maximal clique for Facebook Dataset, and 82 of twitter dataset. we also extracted the maximum clique where in Facebook the maximum clique consist of 15 node and in twitter consist of 7 node

The following figure shows part of maximal clique and the maximum clique graph for Facebook dataset.

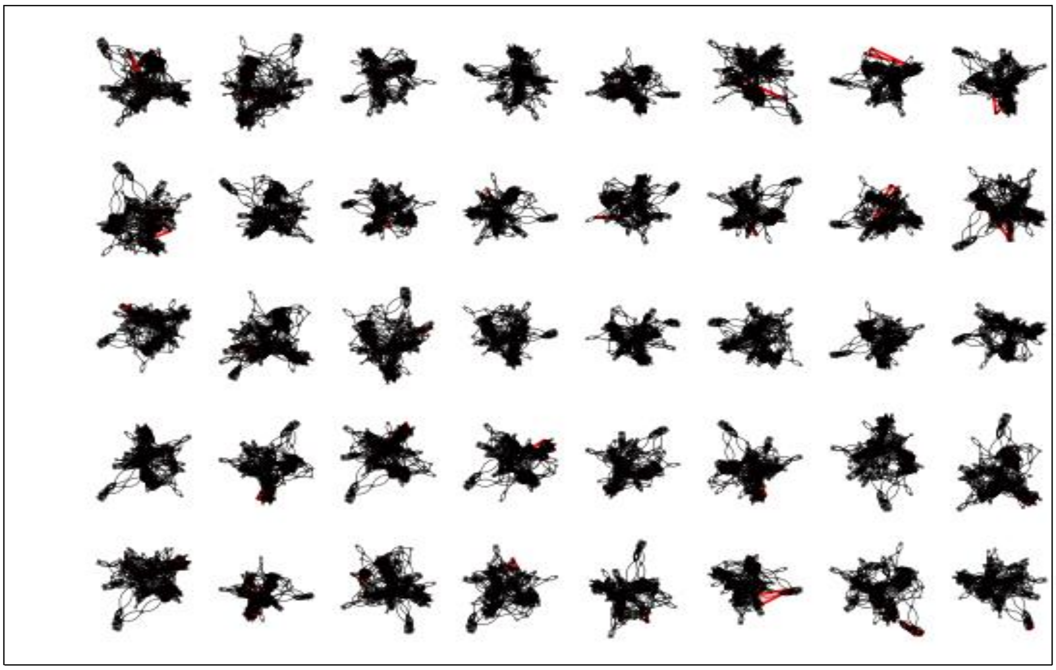


Figure (4.6) graph of Maximal and maximum cliques

4.5 Conventional Leiden algorithm

Here, we initially worked on the original algorithm for comparison, where we entered 250 nodes from Facebook and twitter dataset into the traditional Leiden algorithm and it produced a very large number of communities.

After the stability of the algorithm, it produced approximate 500 communities Where the Leiden algorithm merged similar nodes with each other into one clique when there was an increase in modularity, as communities consisting of two nodes and other more than that appeared, which produced approximate 500 communities as shown in the figure(4.7).

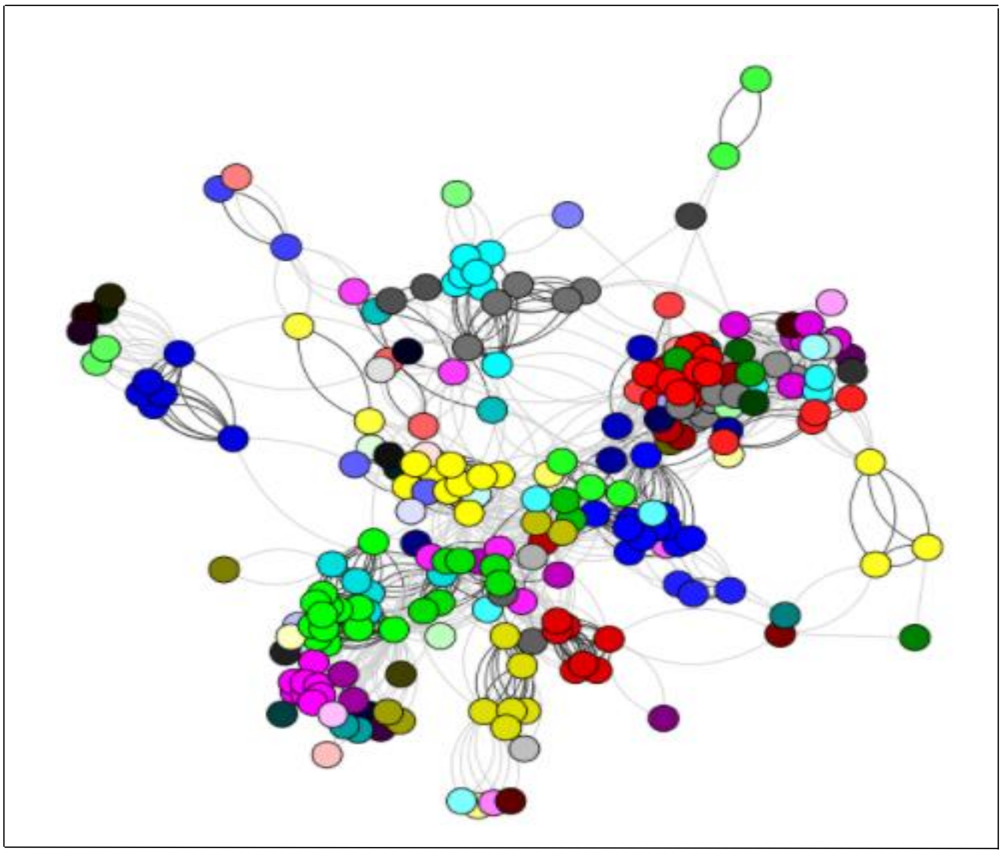


Figure (4.7) the communities of traditional Leiden

Figure 4.8 shows the traditional Leiden algorithm when entering 250 node of twitter dataset As it produced a very large number of communities approximate 312 communities .

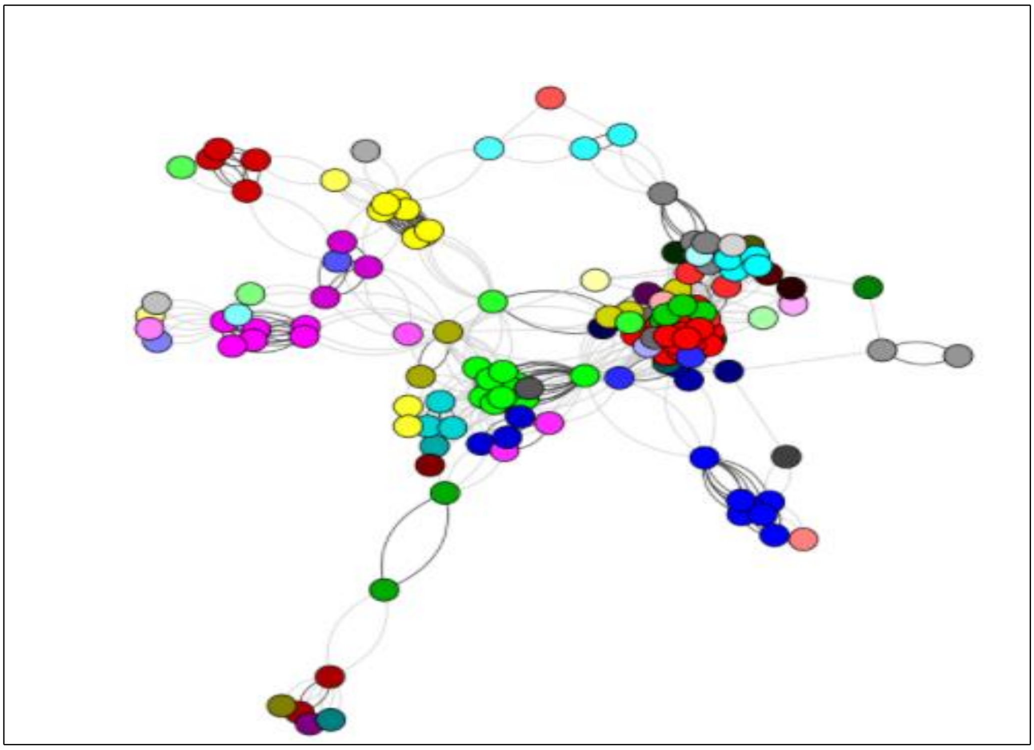


Figure (4.8)Twitter communities in traditional Leiden

4.6 Community Detection

A maximal clique with Leiden is a proposed method to improve the traditional Leiden algorithm. Generally, the key point of our proposed algorithm is looking for the cliques that are inherently found in networks to form the communities.

When entering the maximal clique into the Leiden algorithm, Leiden considers each clique as one node and calculates the modularity for each clique in order to combine it with another clique to form communities.

Here we have combined the maximal clique with the Leiden algorithm ,Where we inserted 102 maximal cliques into the Leiden algorithm that were previously extracted of Facebook dataset .

The following figure shows the communities that were formed after entering 102 maximal cliques into the Leiden algorithm for Facebook Dataset. After introducing the maximal clique into the Leiden algorithm, the Leiden algorithm merged some of the cliques together due to increase of the modularity when placing these cliques together, which means that these cliques belong to the same community, and it also added nodes to some cliques to form communities, where the communities that are formed after the merger consist of 3 nodes as a minimum

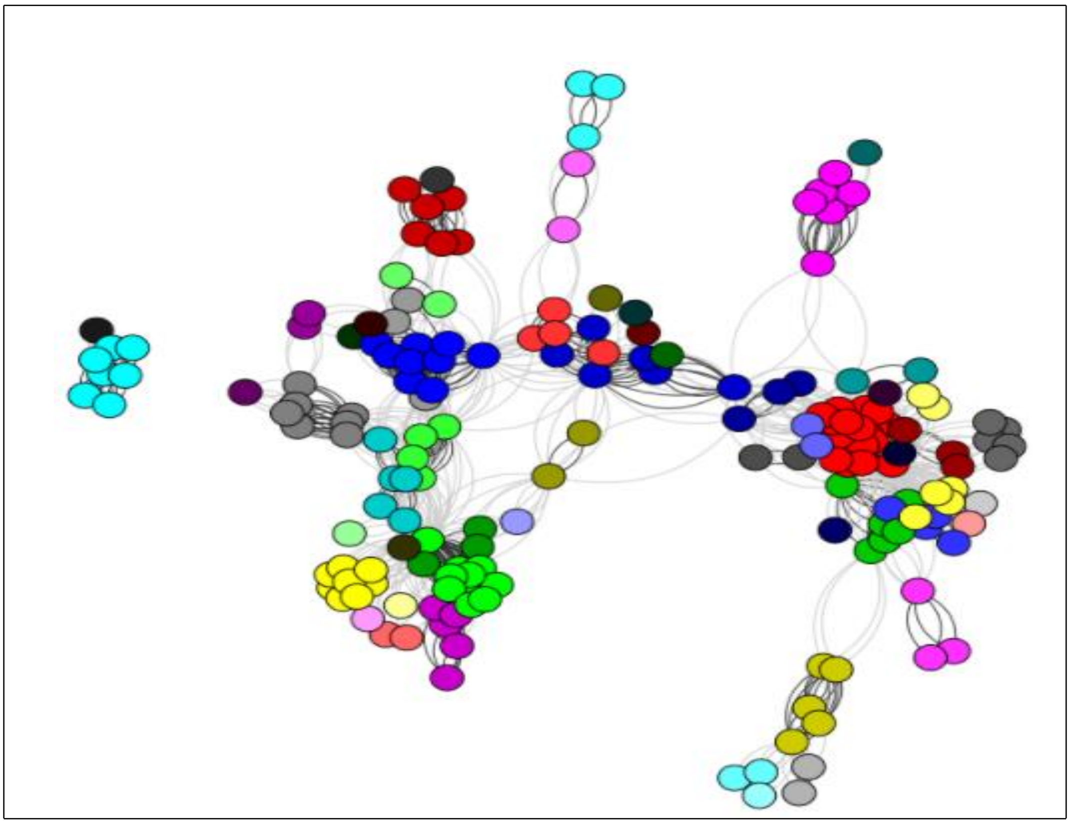


Figure (4.9) maximal clique with Leiden for Facebook dataset

The following figure shows inter 82 maximal clique into the Leiden algorithm which resulted a large number of communities For twitter Dataset

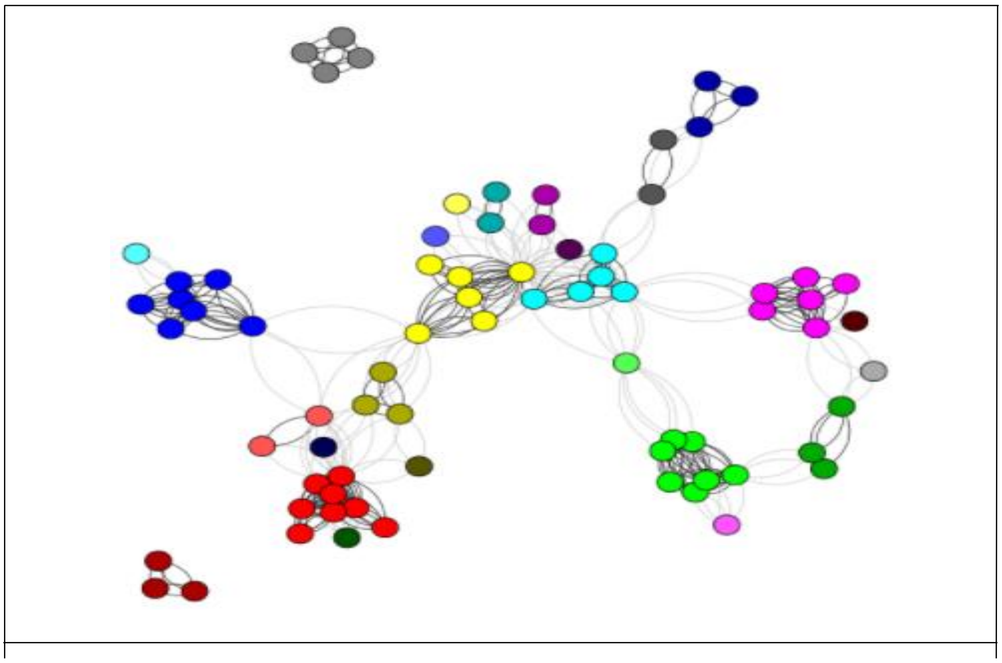


Figure (4.10) maximal clique with Leiden for twitter dataset

4.7 Case study

To ensure the correctness of the work, a small sample was taken from the Facebook Dataset, and this data was implemented according to the previous steps in the previous sessions.

Where we used only 32 nodes, and we extracted the clique from this data, where we got 12 cliques as show in figure 4.12.

```
These are the vaild cliques in the data :
[(1, 5, 22, 27),
 (1, 2, 9, 23),
 (1, 10, 13),
 (1, 14, 21),
 (1, 21, 25),
 (1, 26, 27),
 (28, 29, 31),
 (1, 6, 11, 15, 22),
 (1, 10, 20),
 (1, 2, 11, 17, 18, 23),
 (1, 15, 22, 27),
 (1, 20, 23)]
```

Figure (4.11) cliques of Facebook dataset

Then we extracted the maximum clique from previous 12 cliques, and the number was 11maximal clique.

Then we initially worked on the original algorithm for comparison, where we entered 32 nodes from Facebook dataset into the traditional Leiden algorithm, which produced approximately 98 communities of Facebook dataset.

The following figure shows 32 nodes, which formed approximately 98 communities.

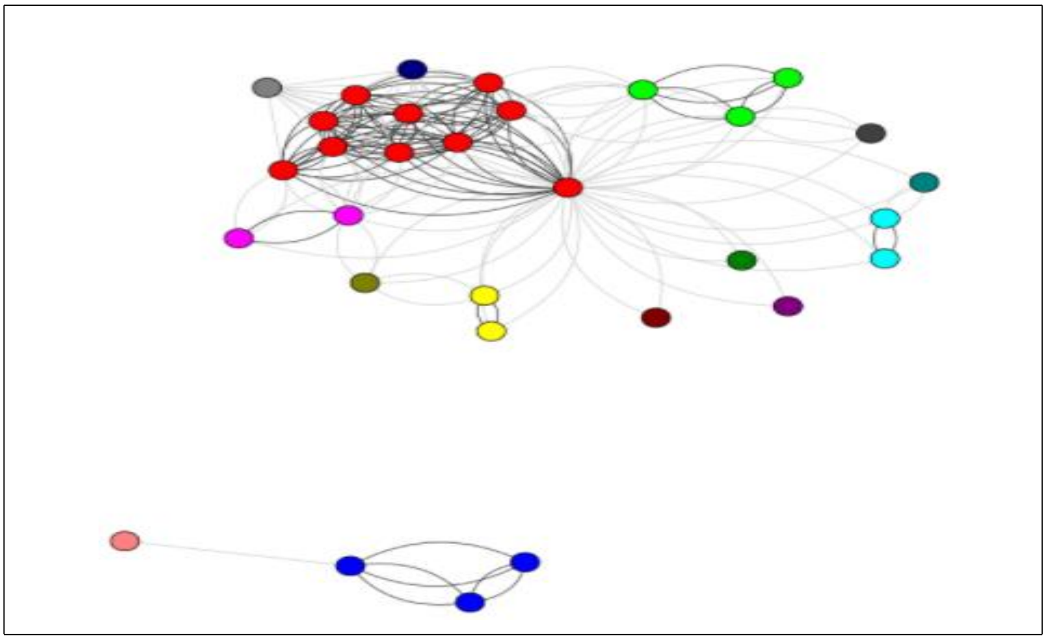


Figure (4.12)Facebook communities in Leiden

The following figure shows the communities that were formed after entering 11 maximal cliques into the Leiden algorithm, which resulted in approximately 42 communities For Facebook Dataset .Where the number of cliques or communities decreased, because the Leiden algorithm here will deal with cliques not a node, as in the traditional case.

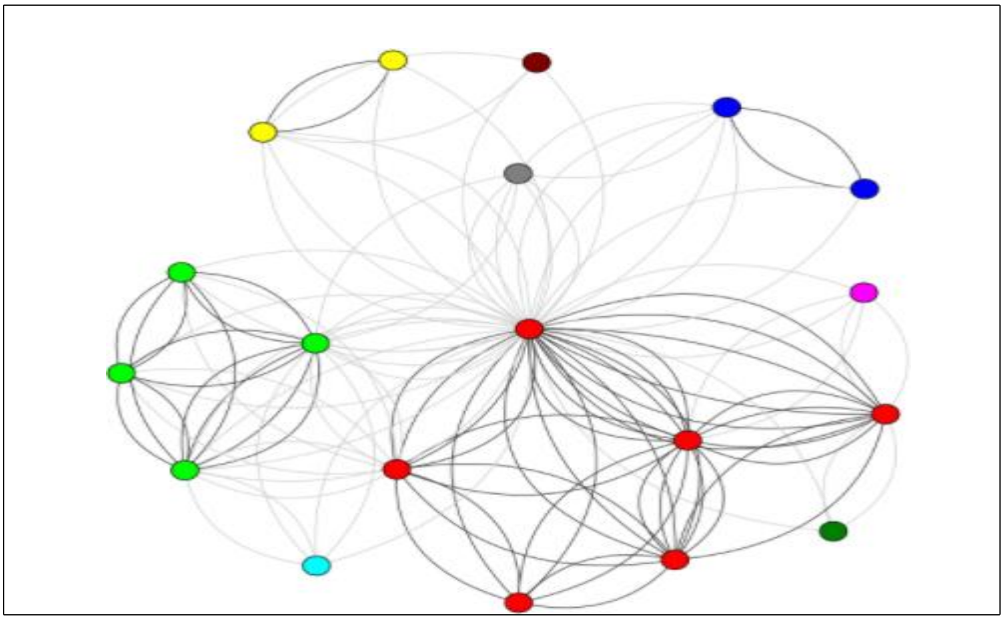


Figure (4.13) maximal clique with Leiden

4.8 Modularity Measurement

A modularity measure is an objective function to evaluate communities' quality. So modularity is the important standard to measure the quality of the Leiden algorithm as explained in equation (2.1) . The results obtained in Figure (4.14) illustrated the modularity measure of Leiden and maximal clique with Leiden algorithms under various datasets. As for the Facebook datasets, the modularity is more than that of the traditional algorithm.

It can be said the results are relatively close to each other .More precisely, the modularity when applied in Facebook of the maximal clique with Leiden algorithm is better than when applied in Twitter. which may be the reason is the structure of the dataset. Indeed, the structure of the Facebook dataset is characterized by largely overlapping cliques, which may affect the modularity. In nature, modularity is very important measures to assess the performance of an algorithm.

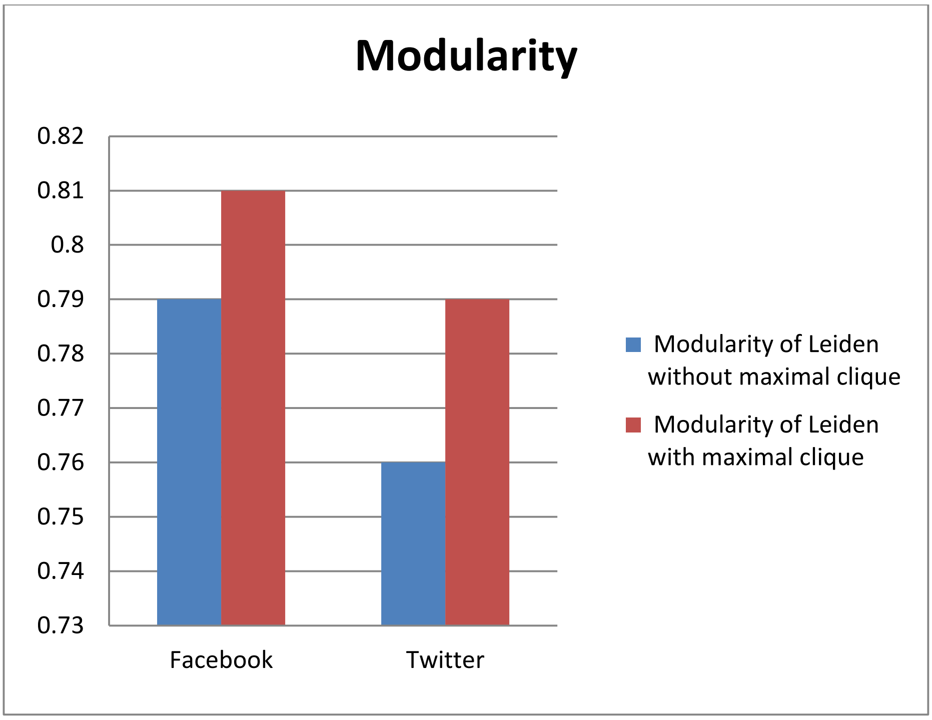


Figure (4.14) Modularity Measure of maximal clique with Leiden and Leiden algorithm

4.9 Results

The flowing table show the modularity value of Facebook and twitter dataset in traditional Leiden algorithm and when merge maximal clique with Leiden algorithm

Table (4.1) Leiden Community Detection Result

Type	Modularity of Leiden without maximal clique	Modularity of Leiden with maximal clique
Facebook	0.79	0.81
Twitter	0.76	0.79

Table (4.2) : The proposed system is compared with recent work

No	Author, Year	Dataset	Algorithm	modularity
1	(Waltman ets al.2019)	Journal	Leiden	0.77
2	(Zhang ets al.2021)	Facebook	Fast Louvain	0.76
3	(Elaf ,2022)	Facebook	clique-Louvain	0.60
4	(Hayder,2022)	Twitter	Leiden	0.79
5	(Fabian, 2021)	Livejournal	Parallel Leiden	0.75
6	(Traag,Waltman and Van,2019) [17]	live journal	Leiden	0.76
#	Our proposed system	Facebook Twitter	Maximal and maximum clique with Leiden	0.81 0.79

CHAPTER FIVE

Conclusions and Future Works

Chapter Five
Conclusions and Future Works

5.1 Conclusions

The research in this thesis proposes a new approach to detect communities in social networks using a combination of the maximal and maximum clique techniques with the Leiden algorithm. The work acknowledges that community detection in social networks is an attractive area of research, but existing algorithms have faced challenges and may not always produce good results in identifying communities.

The key findings and conclusions from the research include:

1. The combination of the maximal and maximum clique techniques with the Leiden algorithm yielded promising results for community detection in social networks.
2. The merger between the maximal and maximum clique with the Leiden algorithm resulted in an improved modularity ratio of 0.81, compared to 0.79 before the merger. This indicates an enhancement in the quality of identified communities.
3. Additionally, the proposed algorithm demonstrated a reduction of 25.91% in the execution time compared to the original Leiden algorithm, specifically when applied to the Facebook dataset. This suggests that the new approach is more efficient for analyzing large-scale datasets.
4. **Maximum Clique Algorithm and Leiden Integration**: The maximum clique algorithm allows us to identify cliques with the

largest number of vertices or edges, representing cohesive subgroups within the network. We then incorporate these maximum cliques into the Leiden algorithm.

5. **Maximal and Maximum Clique Algorithm and Leiden Integration**: Additionally, we leverage the information from maximal cliques, which are cliques that cannot be extended by adding more adjacent vertices, These fully connected substructures provide valuable insights into well-defined communities. The maximum clique algorithm allows us to identify cliques with the largest number of vertices or edges
6. **Enhanced Community Detection**: By integrating the information from both maximum and maximal cliques into the Leiden algorithm, we enhance its ability to identify and assign nodes to communities more efficiently. This comprehensive approach ensures that densely connected subgroups and structurally complete communities are accurately discovered.
7. **Optimized Community Structure**: The utilization of cliques within the Leiden algorithm enables us to achieve a more optimal community structure, leading to a better representation of the underlying relationships and interactions within the social network.
8. **Reduced Effort and Time**: The combination of these approaches not only improves the quality of community detection but also reduces the computational effort and time required to identify meaningful communities in the network.

Overall, the research highlights the potential benefits of combining different algorithms and techniques to enhance community detection in social networks. By using the maximal and maximum clique techniques with the Leiden algorithm, the study achieved improved results in terms of modularity and computational efficiency. This advancement could lead to better insights into the structure and dynamics of communities in social networks and potentially have broader applications in related fields.

5.2 Future Works

There are some potential future works that can be explored to further develop and enhance the proposed community detection model:

1. **Application to Different Social Networks**: The current research focuses on applying the proposed model to Facebook data. To validate its effectiveness and generalizability, future studies can test the model on other social network datasets like Instagram, Twitter, LinkedIn, or any other platform with different characteristics. This will allow researchers to assess how well the model performs on diverse datasets and understand its applicability in various social network contexts.
2. **Incorporation of Additional Features**: The proposed model can be extended by incorporating additional features from the Facebook dataset. Social networks often contain rich information beyond just the network structure, such as user attributes, textual data, timestamps, and more. Integrating such features into the model could potentially lead to more accurate and meaningful community detection results. This avenue of research could open up new possibilities for capturing nuanced community patterns in social networks.

3. **Weighted Network Analysis**: Currently, the proposed model treats the network as an undirected and unweighted graph. However, many real-world networks, including social networks, may have weighted edges that represent the strength or intensity of relationships between nodes. Future work could focus on adapting the model to handle weighted networks and incorporate edge weights into the community detection process. This would enable a more refined analysis, where the strength of connections influences the community structure.

References

[1] L. Burbach, P. Halbach, M. Ziefle, and A. Calero Valdez, “Opinion Formation on the Internet: The Influence of Personality, Network Structure, and Content on Sharing Messages Online,” Front. Artif. Intell., vol. 3, no. July, pp. 1–16, 2020, doi: 10.3389/frai.2020.00045.

[2] N. Hafiene, W. Karoui, and L. Ben Romdhane, “Influential nodes detection in dynamic social networks: A survey,” Expert Syst. Appl., vol. 159, p. 113642, 2020, doi: 10.1016/j.eswa.2020.113642.

[3] C. C. Aggarwal, S. Lin, and P. S. Yu, “On influential node discovery in dynamic social networks,” Proc. 12th SIAM Int. Conf. Data Mining, SDM 2012, pp. 636–647, 2012, doi: 10.1137/1.9781611972825.55.

[4] Y. Yang, Z. Wang, T. Jin, J. Pei, and E. Chen, “Tracking Top-K Influential Vertices in Dynamic Networks,” 2018.

[5] Y. Xie , X. Wang , D.Jiang , R. Xu , High-performance community detection in social networks using a deep transitive autoencoder, 17 April 2019.

[6] S. Papadopoulos, Y. Kompatsiaris, A Vakali, and P. Spyridonos, "Community detection in social media," Data Min. Knowl. Discov., vol. 24, no. 3, pp. 515-554, 2012.

[7] Q. Zhang et al., "Landscape and dynamics of single immune cells in hepatocellular carcinoma," Cell, vol. 179, no. 4, pp. 829-845, 2019.

[8] M. K. Alasadi and H. N. Almamory, “Diffusion model based on shared friends-aware independent cascade,” J. Phys. Conf. Ser., vol. 1294, no. 4, 2019, doi: 10.1088/1742-6596/1294/4/042006.

[9] D. Mohapatra, A. Panda, D. Gouda, and S. S. Sahu, “A Combined Approach for k-Seed Selection Using Modified Independent Cascade Model,” in Computational Intelligence in Pattern Recognition, Springer, 2020, pp. 775–782.

[10] M. Valeri, Social network analysis: organizational implications in tourism

management, International Journal of Organizational Analysis (2020), doi: 10.1108/IJOA-12-2019-1971.

[11] M. Li, S. Lu , L. Zhang, Y. Zhang, and B. Zhang, A Community Detection Method for Social Network Based on Community Embedding, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 8, NO. 2,2021.

[12] G. Bello-Orgaz, J. Hernandez-Castro, and D. Camacho, "Detecting discussion communities on vaccination in twitter," Futur. Gener. Comput. Syst., vol. 66, pp. 125-136, 2017.

[13] E. Jokar and M. Mosleh, "Community detection in social networks based on improved Label Propagation Algorithm and balanced link density." Phys. Lett A, vol. 383, no. 8, pp. 718-727, 2019

[14] V. Trang, L. Wahman, and N. I. Van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Sci Rep, vol. 9, no. 1, pp. 1-12, 2019

[15] E. A. Abbas, "Improving Louvain Algorithm by Leveraging Cliques for Community Detection" , 2020,978-1-7281-5249-3/20.

[16] J. Zhang, J. Fei, X. Song, and J. Feng, “An Improved Louvain Algorithm for Community Detection,” Math. Probl. Eng., vol. 2021, 2021, doi: 10.1155/2021/1485592

[17] H.M ALash, "ENHANCING COMMUNITY DETECTION IN SOCIAL NETWORK USING TOPIC MODELLING AND SENTIMENT ANALYSIS",2022

[18] T. B. N. Hoang, "Topical Community Detection: an Embedding User and Content Similarity Method," 2020.

[19] C. Li, J. Bai, Z. Wenjun, and Y. Xihao, "Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment," Inf. Process Manag, vol. 56, no. 1, pp. 91-109, 2019.

[20] X. You, Y. Ma, and Z. Liu, "A three-stage algorithm on community detection in social networks," Knowledge-Based Syst, vol. 187, p. 104822, 2020

[21] F. Nguyen, "Leiden-Based Parallel Community Detection", At the Department of Informatics Institute of Theoretical Informatics, 2021.

[22] C. D. G. Linhares, J. R. Ponciano, F. S. F. Pereira, L. P. C. Rocha, J. G. S. Paiva, and B. A. N. Travençolo, "Visual analysis for evaluation of community detection algorithms," Multimed Tools Appl., vol. 79, 2020.

[23] A. Dhanasetty, "Enumerating Approximate Maximal Cliques in a distributed framework", B.Tech., Computer Science and Engineering, Gandhi Institute of Technology and Management Visakhapatnam, India, 2021.

[24] N. S. Sattar and, S. Arifuzzaman, "Parallelizing louvain algorithm: Distributed memory challenges," pp. 695–701, 2018, doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00122.

[25] Z. Xu, X. Rui, J. He, Z. Wang, and T. Hadzibeganovic, "Superspreaders and superblockers based community evolution tracking in dynamic social networks," Knowledge-Based Syst., vol. 192, p. 105377, 2020, doi: 10.1016/j.knosys.2019.105377.

[26] M. Valeri, R. Baggio, "Social network analysis: organizational implications in tourism management", (2020), doi: 10.1108/IJOA-12-2019-1971.

[27] B. Chang and E. Chen, "Study on Information Diffusion Analysis in Social Networks and Its Applications," vol. 15, no. August, pp. 377– 401, 2018, doi: 10.1007/s11633-018-1124- 0.

[28] K . Nallusamy , K .S. Easwarakumar , "CGraM: Enhanced Algorithm for Community Detection in Social Networks" , DOI:10.32604/iasc.2022.020189 ,2021.

[29] M. Valeri, "Social network analysis: organizational implications in tourism management", International Journal of Organizational Analysis (2020), doi: 10.1108/IJOA-12-2019-1971.

[30] H. Fani, E. Bagheri , "Community detection in social networks ",Laboratory for Systems, Software, and Semantics (LS3) Ryerson University, Toronto, ON, Canada, ,17 March 2020

[31] A. Bouyer, "Community Detection in Social Networks Using Affinity Propagation with Adaptive Similarity Matrix", 2020 DOI: 10.1089/big.2019.0143.

[32] D. Choudhury , A. Paul, "COMMUNITY DETECTION IN SOCIAL NETWORKS: AN OVERVIEW" Assam University, Silchar, , Dec-2019, Available.

[33] D. Camacho, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E Cambria, "The four dimensions of social network analysis: An overview of research methods", applications and software tools," Inf Fusion, vol. 63, pp. 88-120, 2020.

[34] P. Bedi and C. Sharma, "Community detection in social networks," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 6, no. 3, pp. 115-135, 2016.

[35] N. Hafiene, W. Karoui, and L. Ben Romdhane, "Influential nodes detection in dynamic social networks: A survey," Expert Syst. Appl., vol. 159, p. 113642, 2020, doi: 10.1016/j.eswa.2020.113642.

[36] C. C. Aggarwal, S. Lin, and P. S. Yu, "On influential node discovery in dynamic social networks," Proc. 12th SIAM Int. Conf. Data Mining, SDM 2012, pp. 636–647, 2012, doi: 10.1137/1.9781611972825.55.

[37] Y. Yang, Z. Wang, T. Jin, J. Pei, and E. Chen, "Tracking Top-K Influential Vertices in Dynamic Networks," 2018.

[38] D. T. Schroeder, K. Pogorelov, and J. Langguth, "Fact. a framework for analysis and capture of twitter graphs," in 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2019, pp. 134-141.

[39] J. Shang, L. Liu, X. Li, F. Xie, and C. Wu, "Targeted revision : A learning-based approach for incremental community detection in dynamic networks," Physica A, vol. 443, pp. 70–85, 2016, doi: 10.1016/j.physa.2015.09.072.

[40] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels, "Scalable community detection with the Louvain algorithm," in 2015 IEEE International Parallel and Distributed Processing Symposium, 2015, pp. 28-37.

[41] X. Yuan, R. J. Schuchard, and A. T. Crooks, "Examining emergent communities and social bots within the polarized online vaccination debate in Twitter," Soc. media+ Soc., vol. 5, no. 3, p. 2056305119865465, 2019.

[42] J. Han, W. Li, L. Zhao, Z. Su, Y. Zou, and W. Deng, "Community detection in dynamic networks via adaptive label propagation," PLoS One, vol. 12, no. 11, pp. 1–16, 2017, doi: 10.1371/journal.pone.0188655.

[43] S. Taheri and A. Bouyer, "Community Detection in Social Networks Using Affinity Propagation with Adaptive Similarity Matrix", 2020, DOI: 10.1089/big.2019.0143.

[44] D. Choudhury , A. Paul, “COMMUNITY DETECTION IN SOCIAL NETWORKS: AN OVERVIEW”, Triguna Sen School of Technology, Silchar,2013 eISSN: 2319-1163.

[45] H. Fani , E. Bagheri, “Community detection in social networks”, Laboratory for Systems, Software, and Semantics (LS3) Ryerson University, Toronto, ON, Canada,2017 DOI: 10.1142/S2425038416300019.

[46] P. Bedi , C. Sharma , “Community detection in social networks”, University of Delhi, DOI: 10.1002/widm.1178, 2016. ·

[47] S. K. Gupta ,D. P. Singh, J. Choudhary,” A review of clique-based overlapping community detection algorithms”, Maulana Azad National Institute of Technology, doi.org/10.1007/s10115-022-01704-6,2022.

[48] Z. Liu, Y. Ma, “A Divide and Agglomerate Algorithm for Community Detection in Social Networks”, Shandong, 250014, P. R. China,2019 DOI: 10.1016/j.physa.2019.04.061.

[49] X. Li, X. Cheng, S. Su, and C. Sun, “Community-based seeds selection algorithm for location aware influence maximization,” Neurocomputing, vol. 275, pp. 1601–1613, 2018, doi: 10.1016/j.neucom.2017.10.007.

[50] T M. Consortium, "A single cell transcriptomic atlas characterizes aging tissues in the mouse," Nature, vol. 583, no. 7817, p. 590, 2020.

[51] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” Phys. Rev. E, vol. 69, no. 2, p. 26113, 2004.

[52] X. Bai, P. Yang, and X. Shi, "An overlapping community detection algorithm based on density peaks," Neurocomputing, vol. 226, pp. 7-15, 2017

[53] A. Song, M. Li, X. Ding, W. Cao and K. Pu, “Community Detection Using Discrete Bat Algorithm, School of Computer Engineering and Science”, Shanghai, 200444 China,2016 IJCS_43_1_05.

[54] H. Zhou, J. Li, J. Li, F. Zhang, and Y. Cui, "A graph clustering method for community detection in complex networks”, Phys A Stat. Mech. Its Appl vol. 469, pp. 551-562, 2017.

[55] V. D. Blondel, J-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat Mech theory Exp, vol 2008, no. 10, p. P10008, 2008.

[56] VA Trang, L. Wahman, and N. I. Van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Sci Rep, vol. 9, no. 1, pp. 1-12, 2019

[57] S. J. Blair, Y. Bi, and M. D. Mulvenna, "Aggregated topic models for increasing social media topic coherence," Appl. Intell, vol. 50, no. 1, pp. 138-156, 2020.

[58] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring topic coherence over many models and many topics," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, pp. 952- 961.

[59] C. I. Wood, “Clique Generalizations and Related Problems”, RICE UNIVERSITY, 2015.

[60] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, , and M.-S. Alouini, “A Tutorial on Clique Problems in Communications and Signal Processing”, the Department of Electrical Engineering, California Institute of Technology, Pasadena, 2020 arXiv:1808.07102v4.

[61] S. K. Gupta, D. P. Singh and J. Choudhary, “A review of clique-based overlapping community detection algorithms”, Maulana Azad National Institute of Technology, Knowledge and Information Systems (2022) 64:2023–2058.

[62] K. Nallusamy and K. S. Easwarakumar, “CGraM: Enhanced Algorithm for Community Detection in Social Networks”, College of Engineering, Guindy, Anna University, Chennai, Tamilnadu, India, 17 June 2021.

[63] L. Chang, “Efficient Maximum Clique Computation over Large Sparse”, The University of Sydney, 2019, doi.org/10.1145/3292500.3330986.

[64] L. Chang, “Efficient Maximum Clique Computation over Large Sparse Graphs”, The University of Sydney, Australia, July 2019 doi.org/10.1145/3292500.3330986.

[65] L. F. Litzka, “Analysis of a Maximal Clique Finding Algorithm with respect to Runtime and Eectiveness in High Dimensional Data”, College of Mathematics, Informatics and Statistics ,Ludwig Maximilian-University of Munich, 28.08.2019.

[66] X. Li, X. Cheng, S. Su, and C. Sun, “Community-based seeds selection algorithm for location aware influence maximization,” Neurocomputing, vol. 275, pp. 1601–1613, 2018, doi: 10.1016/j.neucom.2017.10.007.

[67] T. B. N. Hoang, "Topical Community Detection: an Embedding User and Content Similarity Method," 2020.

[68] A. Culotta, "Towards detecting influenza epidemics by analyzing Twitter messages," in Proceedings of the first workshop on social media analytics. 2010, pp 115-122

الخلاصة

في الوقت الحاضر ، توسعت الشبكات الاجتماعية بسرعة. يستخدم الناس الشبكات الاجتماعية لمشاركة آرائهم وأفكارهم ومشاعرهم. يميل الناس إلى بناء مجموعات قائمة على العلاقات مع الآخرين. أحد المجالات الأكثر أهمية التي تساعد في فهم وتحليل بنية الشبكات الضخمة والمعقدة ، مثل الشبكات الاجتماعية والشبكات التعاونية ، هو اكتشاف المجتمعات.

تم تقديم اكتشاف المجتمع كطريقة للعثور على مجتمعات في الشبكات الاجتماعية عن طريق تقسيم الشبكة إلى مجموعات من المستخدمين المتصلين بشكل وثيق. العناصر المهمة للبحث هي استخراج المعلومات ذات الصلة من هذه الشبكات. الهدف من اكتشاف المجتمع هو تقليل الرسم البياني الناتج عن التطبيق إلى مجتمعات أصغر ذات عقد قابلة للمقارنة. يعد عد المجموعات في شبكة أكبر مشكلة أساسية في نظرية الرسم البياني. تركز الدراسة الحالية على شبكة هيكلية تعتمد فقط على الروابط بين المستخدمين لاكتشاف المجتمعات.

تم تطوير النموذج المقترح لتعزيز طريقة اكتشاف مجتمع Leiden. قمنا بدمج خوارزمية ليدن مع خوارزميات maximal and maximum clique ، حيث تم حساب المجموعات أولاً ، ثم تم استخراج maximal and maximum clique من هذه المجموعات ، ثم تم إدخال maximal and maximum clique الى خوارزمية ليدن. أخيرًا ، يمكن للنموذج المقترح أن يعزز وظيفة الجودة في طريقة الكشف عن مجتمع Leiden. تم إجراء العديد من التجارب على فيسبوك و تويتر. أظهرت النتائج تحسنًا في جودة ليدن للنمطية ، حيث زادت النمطية في مجموعة بيانات فيسبوك من 0.79 إلى 0.81 وأيضًا في تويتر زادت من 0.76 إلى 0.79

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل
كلية تكنولوجيا المعلومات
قسم شبكات المعلومات

خوارزميه ليدن المحسنة بأستخدام الزمر

رسالة

مقدمة إلى مجلس كلية تكنولوجيا المعلومات في جامعة بابل كجزء من متطلبات نيل درجة الماجستير في تكنولوجيا المعلومات – شبكات المعلومات

مقدمة من قبل

بنين علي كاظم محمد

باشراف

أ.د. غيداء عبد الحسين الملا

١٤٤٥ هـ　　　　٢٠٢٣م