Thesis submitted to the

UNIVERSITY OF MOHAMED BOUDIAF – MSILA

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

In partial fulfillment of the requirements for the degree of

Master in Computer Science

OPTION: Informatique Décisionnelle et Optimisation (IDO)

By

Chami, Wiam **Radja, Imane**

A Comparative Study of Overlapping Community Detection Algorithms in Social Networks

Under the supervision of

Bilal Lounnas

Composition of the jury

Dr.BOUNIF Mohamed	University of Msila	President
Dr.LOUNNAS Bilal	University of Msila	Supervisor
Dr.LOUCIF Hemza	University of Msila	Examinator

June, 2025

Dedication

“I dedicate my graduation

To the one whose name I carry with pride... My support in this life, and my companion in prayers, To the one who taught me that willpower makes the impossible possible, And that patience is the gateway to all that is beautiful ***my dear father****.*

And to the source of sincere prayers, my paradise on earth, The fountain of tenderness and the pillar of safety ***my beloved mother****.*

To my dear sisters: ***Ines*** *and* ***Khadija****, You were with me every step of the way — the joy in times of exhaustion, the strength when hope faded.*

To my little brother ***Ilyes****, the joy of our home and the light of my heart, Your smile lit up my darkest days.*

To ***my friends*** *who were like a second family, Those who shared with me laughter and worry, success and hardship,*

To ***my esteemed professors****, who instilled in me a love for knowledge, And opened before me the doors of understanding, critical thinking, and discovery,*

And to everyone who supported me with a word, a prayer, or a smile,

And finally... Praise be to Allah, by whose grace good deeds are completed. All thanks to Him for what He has given, and all gratitude for His blessings. He is the Guide and the Helper, and from Him alone I draw strength and determination for every step ahead.”

chami wiam

Dedication

To my Self...

Who never give up, who stayed up countless nights and endured the hardship in recognition of true determination, and in loyalty to every moment of patience and perseverance.

To those who, after God's grace, were the light that guided my path… To those who instilled in me values and principles, and were my support and strength…

To my dear father, the source of my strength, To my kind mother, the fountain of love and sincere prayers… To you both, all my love and deepest gratitude.

To my brothers, who always encouraged and supported me.

To my friends, who were with me in hard and happy times, and supported me with kind words and prayers.

To my esteemed teachers, who instilled in me the love of knowledge and diligence.

Radja Imane

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful.

All praise and thanks be to Allah, first and last, outwardly and inwardly, for guiding us, granting us success, and easing our path in both knowledge and work. To Him belongs all gratitude and praise, as is worthy of His majestic countenance and supreme authority.

We extend our sincerest thanks, deepest appreciation, and profound gratitude to everyone who contributed, whether directly or indirectly, to the completion of this thesis.

First and foremost, we express our heartfelt gratitude and deep appreciation to our esteemed professor and supervisor, **Dr.Bilal Lounnas**, for his constant support, valuable guidance, continued encouragement, and great patience throughout this research. His insightful observations and wise direction illuminated our path at every step of the way. May Allah reward him abundantly and bless his knowledge and efforts.

We would also like to extend special thanks to our colleagues, Houssam Bada and Mohamed Hamrit, who never hesitated to lend a helping hand. They have our full respect and appreciation.

We are equally grateful to my dear friend Fadwa Yahi, who kindly offered me her personal computer when mine broke down. Her generous gesture came at a critical moment and greatly helped me to carry on with my work. I offer her my heartfelt thanks and appreciation.

We also express our gratitude to the Dean of the Faculty of Mathematics and Computer Science, Professor **Ibrahim Nouiri**, and to all the faculty members at Mohamed Boudiaf University of M'sila, for the support and facilities they provided throughout the preparation of this thesis, especially during difficult times. Their support was essential to the success of this work.

Finally, we extend our warmest thanks and deepest appreciation to our families, friends, and everyone who stood by us, encouraged us, and provided moral support throughout this journey. Without them, this achievement would not have been possible.

We hope this research will contribute to enriching knowledge and that students will benefit from it in their future studies.

All praise is due to Allah, Lord of the Worlds.

Contents

List of Tables

List of Figures

General introduction

Social networks are one of the most often used models in a variety of disciplines, including economics, computer science, and sociology, to comprehend the relationships and interactions between people and groups. These networks show a complicated structure made up of nodes, which stand for people or things, and edges, which show connections or interactions between them. A key feature of these networks is the existence of communities, which are subsets of nodes having more connections amongst each other than with the rest of the network.

Detecting these communities helps in the comprehension of the network's internal structure and supports a variety of useful applications, including the analysis of user behavior, product recommendations, refining marketing campaigns, and even researching the dissemination of information or diseases.

- **The importance of overlapping community detection**

  There may be overlapping communities in various social networks since a person may be a member of multiple communities at the same time. Because algorithms must be created to precisely and effectively handle such overlaps, this intricacy makes community detection a major task. Finding overlapping communities gives a more accurate picture of interpersonal relationships and a more realistic representation of social systems.

- **Research motives**

  With the rapid expansion in the use of social networks and the diversity of their applications, it has become essential to develop effective methods for detecting overlapping communities. This study aims to compare and analyze various approaches and algorithms used in this field, with a focus on evaluating their performance in terms of accuracy, efficiency, and their ability to handle community overlap, by using metrics such as ONMI and Extended Modularity.

  We are interested in understanding the strengths and weaknesses of each method, enabling us to provide practical recommendations for selecting the most suitable one based on the nature of the network and the intended application.

- **Research objectives**

  - Studying the fundamental concepts related to overlapping communities in social networks.
  - Reviewing and analyzing overlapping community detection algorithms, with a focus on modern approaches.
  - Conducting a comprehensive comparison of these algorithms using standardized evaluation metrics.

- **Organization of the research**

  This research is structured into three main chapters, each focusing on critical aspects related to social networks and overlapping community detection.

  - The first chapter, **Literature Review** Provides a comprehensive review of social networks and their analysis. It begins with definitions, history, and types of social networks, followed by an introduction to graph theory concepts relevant to social networks. The chapter also covers key metrics used in social network analysis such as centrality, clustering coefficient, and network density. This chapter establishes the theoretical foundation for understanding community structures within networks.
  - The second chapter **Overlapping Community Detection** Focuses on overlapping community detection methods. It starts by defining communities and their significance in social networks, then reviews traditional and overlapping community detection algorithms. This chapter also discusses objectives, challenges, classifications, and applications of overlapping community detection, providing a detailed overview of current approaches and methodologies.
  - The third chapter **Comparison and Results** Presents the evaluation framework, including metrics like modularity and normalized mutual information, used to assess the performance of overlapping community detection algorithms. It describes the datasets, both real and synthetic, the experimental setup, and the algorithms compared (such as SLPA, CPM, BigCLAM, DEMON, and Ego-Splitter). The chapter concludes with the presentation and analysis of experimental results, highlighting the strengths and weaknesses of each approach.

By exploring the fundamental concepts, specific and specialized algorithms, and the practical implementation of overlapping community detection in social networks, this research aims to contribute to the advancement of this field through a comparative study of the most prominent methods found in the literature.

Chapter 1: Literature Review

1.1 Introduction

In the digital age, social networks have evolved into powerful platforms that not only connect individuals but also shape communication, commerce, and society at large. From their early inception as simple tools for personal connection to their current role as complex, multidimensional systems, social networks have transformed the way we interact, exchange information, and engage with the world around us. Understanding the structure, dynamics, and impact of social networks is crucial, as they influence everything from marketing strategies to political movements. This literature review delves into the key concepts of social networks, exploring their definition, history, various types, and their importance in modern society, with a focus on the role they play in communication, business, and cultural transformation. Additionally, it will cover the analysis of social networks through the lens of graph theory, which provides valuable insights into the relationships and behaviors that define these networks.

1.2 Social networks

Social networks are complex structures consisting of individuals or entities connected through various relationships, such as friendship, collaboration, or online interaction. The study of social networks has become increasingly important in various fields, including sociology, marketing, and data science.

1.2.1 Definition

A social network is a structure consisting of individuals or entities (called nodes) connected by links (called edges) that represent various types of interactions such as friendships, professional collaborations, social influences, or business exchanges. These networks reflect the relationships between individuals within a society and help understand how these connections influence social structures and the spread of information or influences. Social networks can be directed (when relationships are not mutual) or undirected (when relationships are mutual). The presence of hubs-individuals with greater influence within the network-plays a significant role in shaping interactions, and both strong and weak ties affect the flow of social processes such as information dissemination, idea spread, or even disease transmission. Through mathematical analysis methods, social networks provide insights into the dynamics of social interactions, helping us study how individuals and communities impact each other in various contexts. [1][2]

1.2.2 Historic

Social media began in the late 1990s with the launch of early platforms like SixDegrees.com (1997), which allowed users to create profiles and connect with friends. Although it didn't achieve significant success and was short-lived, it laid the foundation for the concept of social networks. Later, platforms such as Friendster (2002) and My Space (2003) emerged in the early 2000s. Friendster was one of the first networks aimed at connecting friends online, while My Space gained massive popularity by focusing on music and entertainment, but eventually failed due to the rise of other platforms. Then came Facebook in 2004, which completely changed the game. Initially limited to Harvard University, it quickly expanded to become the largest social network in the world. This shift made social media an essential part of our daily lives, turning it into more than just a communication tool but a platform for both social interaction and commercial engagement. Around the same time, platforms like Twitter (2006) emerged, focusing on real-time communication through short tweets. Over time, social media transformed from just a space for connecting with friends to massive commercial tools, where companies began collecting personal data from users to target them with advertisements. Furthermore, social media increasingly influenced political life, whether through election campaigns or social movements. By 2013, social media had become a dominant force in daily life, influencing everything from politics to business to personal interactions. Platforms like Instagram (launched in 2010) and Snapchat (2011) had also gained popularity, adding new dimensions to online interaction. Social media networks were now deeply ingrained in the cultural fabric of society, continuing to evolve and shape communication in profound ways. [3]

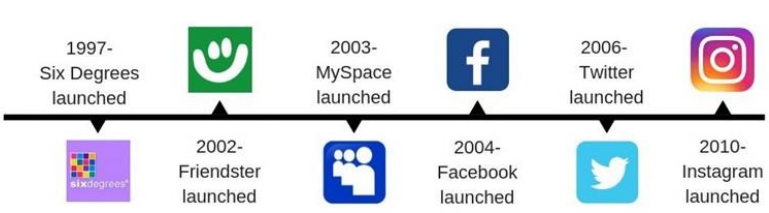

Figure 1.1: Historic of social networks
[4]

1.2.3 Types

Social networks vary in their objectives and functionality, and they differ in how they allow users to interact and share content. However, there are some common characteristics that unify them, such as user profiles, interaction between individuals, and content sharing. Based

on the primary purpose of these networks, they can be categorized into different types. [2] [5]The following table outlines the main types of social networks, with brief definitions and examples for each type:

Type of Social Network	**Definition**	**Examples**
Personal Social Networks	Used for interaction between individuals and sharing daily life and personal content.	Facebook, Instagram, Snapchat
Professional Social Networks	Aimed at building professional relationships and connecting individuals with companies and employers.	LinkedIn, ResearchGate, Xing
Media Sharing Networks	Used for sharing visual content such as photos, videos, and music.	YouTube, TikTok, Pinterest, Flickr
Microblogging Networks	Allows posting short and quick content such as news and instant updates.	Twitter (X), Tumblr, Mastodon
Interest-Based Networks	Connects users based on common interests rather than personal relationships	Reddit, Quora, Stack Overflow
Market Networks	Focused on online selling and e-commerce, connecting buyers and sellers.	Amazon, eBay, Etsy
Collaboration and Open Projects Networks	Used for collaboration on projects, sharing knowledge and resources.	GitHub, Wikipedia, OpenAI Community

Table 1.1: Type of social network [2] [5]

1.2.4 Importance of social networks

Social networks have become an essential part of our daily lives, providing a platform that enables individuals to easily connect with others and strengthen social relationships. Through these networks, people can interact with a wider circle of friends and family, even across

geographical boundaries, making communication easier and enhancing social bonds. On the business front, social networks have become a key tool in marketing and advertising strategies. By collecting massive amounts of user data, companies can target their audience accurately through personalized ads, increasing the effectiveness of campaigns and achieving better results. Moreover, social networks play an important role in the political and social spheres, as they have become a primary platform for expressing opinions and participating in social movements. These networks make it easier to spread political messages and raise awareness about social issues. In terms of digital culture, social networks have contributed to shaping individuals' cultural identities. Platforms like Facebook and Instagram provide space for people to express them- selves and share their thoughts and images, reflecting a significant impact on personal culture. Finally, social networks facilitate commercial and economic innovation, as they have become an integral part of e-commerce. They allow businesses to build direct relationships with cus- tomers and offer personalized services, boosting opportunities for success and growth.[3]

1.3 Graph and social networks analysis

Social Network Analysis (SNA) is a field that focuses on studying the structural patterns of relationships between entities within networks. This analysis is based on graph theory to represent and study the relationships between individuals or different entities, helping to understand how network structure influences individual and group behavior. It is used in analyzing social networks, tracking information flow, and studying collective decision-making processes. Its applications extend to various fields such as sociology, biology, and computer science, contributing to the identification of communication patterns, interaction dynamics, and the evolution of communities and complex systems.[1]

1.3.1 Graph theory

Graph theory is a branch of mathematics that focuses on studying mathematical structures composed of nodes and edges, serving as an analytical model for representing complex relationships between different entities. This theory forms the foundation for many algorithms used in social network analysis, transportation networks, the internet, and other complex systems. By examining the structural properties of networks, graph theory provides deep insights into how network connections influence individual and group interactions, making it a crucial tool in various fields such as sociology, data science, and computer science.[2]

1.3.1.1 Node

A node is the basic unit of a network, representing entities such as individuals, devices, or web pages. In social networks, nodes may be people or groups and serve as the core of network analysis. They represent actors whose relationships reveal the overall network structure. Nodes can also carry attributes like weight, indicating their importance such as follower count on social media.[2]

Figure 1.2: Node representation [6]

1.3.1.2 Edge

An edge connects two nodes in a graph, representing their relationship or interaction, and is key to shaping the network's structure.[2]

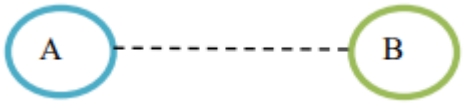

Figure 1.3: Representation of a link between two actors [6]

Edges vary in type depending on the nature of the relationship they represent. they represent 2 types:

Edge Type	Description	Example
Directed Edge	Represents a relationship with a specific direction.	One person following another on social media without reciprocity.
Undirected Edge	Indicates a mutual relationship between two nodes.	Friendships on Facebook, where the connection is bidirectional.

Table 1.2: Edge type [2]

1.3.1.3 Definition of graph

- **Theoretically:**

  A graph is a diagram made up of a set of points connected by binary relations between them. The points are called the vertices of the graph, and the arrows are called the arcs or edges of the graph.[7]

- **Formally:**

  1. A graph is called the pair $G(X, U)$ such that:

  - $X = \{x_1, x_2, \ldots, x_n\}$ is the set of vertices of the graph.
  - $U = \{u_1, u_2, \ldots, u_m\}$ is the set of edges (or arcs) of the graph.
  - $U \subseteq X \times X$, the set of edges is a subset of the Cartesian product of the set of vertices with itself.[8]

  2. A graph $G = (X, U)$ is a mathematical structure consisting of:

  - The number of vertices is denoted by n, where $n = |X|$, and the number of edges is denoted by m, where $m = |U|$.
  - The number of vertices n defines the *order* of the graph.
  - The number of edges m defines the *size* of the graph.
  - Each edge $u \in U$ connects two vertices $\{i, j\}$, and may define a direction (from i to j, or vice versa).
  - If the connections between vertex pairs are bidirectional, the graph is called an *undirected graph*.
  - If the connections are unidirectional, the graph is called a *directed graph*.[9]

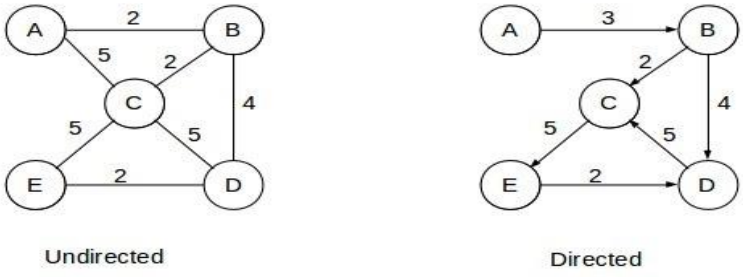

Figure 1.4: Directed graph and undirected graph

1.3.1.5 Types of graphs

The types of graphs used to represent social networks vary according to the characteristics of the relationships between the nodes. Below are the main types:

1. **Symmetric graph**

   - A graph that exhibits a high degree of symmetry, where any pair of edges can be mapped to another through a transformation that preserves the graph's structure.[8]
   - Formally, a graph $G(X, U)$ is said to be symmetric if:

$$\forall x, y \in X, (x, y) \in U \Rightarrow (y, x) \in U$$

   - A symmetric graph is typically represented without directed edges, and the connections are referred to as edges rather than arcs.[8]

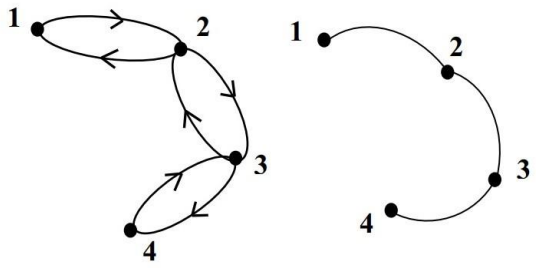

Figure 1.5: Symmetric graph [8]

2. **Complete graph**

- A simple undirected graph in which there is exactly one edge connecting every pair of distinct vertices.[8]
- A graph $G(X, U)$ is said to be complete if and only if:[8]

$$\forall x, y \in X, (x, y) \in U \text{ and } (y, x) \in U$$

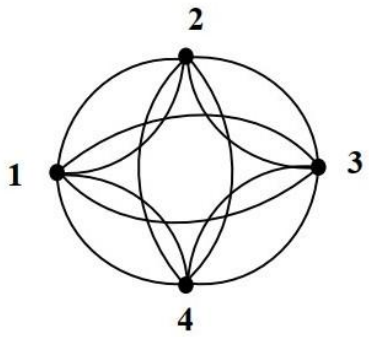

Figure 1.6: Complete graph [8]

3. **Simple graph**

   - A graph that has no multiple edges between the same pair of vertices and does not contain any loops (edges connecting a vertex to itself).[10]
   - A graph $G(X, U)$ is said to be simple if and only if:
     - It contains no loops.
     - For all $x, y \in X$, there exists at most one edge $u = (x, y) \in U$.[8]

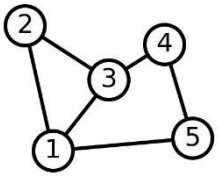

Figure 1.7: Simple graph

4. **Empty graph**

   - A graph that contains no vertices and no edges. [11]
   - A graph $G(X, U)$ is said to be empty if:[8]

$$X = \emptyset, \quad U = \emptyset$$

5. **Trivial graph**

   - A graph that consists of a single vertex and no edges. [12]
   - A graph $G(X, U)$ is said to be trivial if it has vertices but no edges: [8]

$$U = \emptyset$$

6. **Reflexive graph**

   - A graph in which each vertex has a self-loop, meaning every node is connected to itself. [13]
   - A graph $G(X, U)$ is said to be reflexive if: [8]

$$\forall x \in X, (x, x) \in U$$

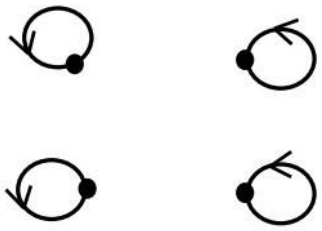

Figure 1.8: Reflexive graph [8]

7. **Antisymmetric graph**

   - A graph in which if there is a relation from node A to node B, there cannot be a relation from B to A, unless A = B. [14]
   - A graph $G(X, U)$ is said to be antisymmetric if: [8]

$$\forall x, y \in X, (x, y) \in U \Rightarrow (y, x) \notin U \text{ (unless } x = y)$$

8. **Transitive graph**

   - A graph in which if there is an edge from A to B and from B to C, then there is also an edge from A to C. [15]
   - A graph $G(X, U)$ is said to be transitive if: [8]

$$\forall x, y, z \in X, (x, y) \in U \text{ and } (y, z) \in U \Rightarrow (x, z) \in U$$

9. **Multigraph**

- A graph that allows multiple edges between the same pair of vertices and may include loops. [16]
- If G is a directed graph, and x and y are two vertices connected by two arcs u_1 and u_2 in the same direction, then G is called a multigraph, and the arc (x, y) is called a multiple arc. [8]

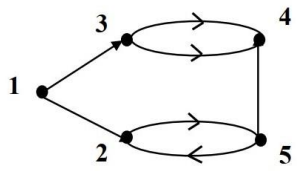

Figure 1.9: Multigraph graph [8]

1.3.1.6 Geodesic distance

The geodesic distance refers to the shortest possible path between two nodes in a graph, measured by the number of edges that must be traversed. For example, in **Figure 1.10**: [6]

- The geodesic distance between node A and node C is **2**.
- The geodesic distance between node A and node B is **1**.

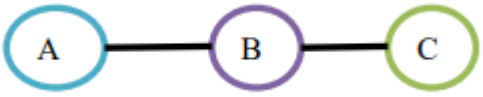

Figure 1.10: Representation of geodesic distance [6]

1.3.2 Social networks analysis

In the following sections, we will present key social network analysis metrics through the lens of graph theory.

1.3.2.1 Diameter

The diameter of a network is the longest shortest path between any two nodes. It measures how far apart nodes can be, helps assess communication efficiency, and reveals the overall spread of the network. [1][2]

It is defined by the formula:[6]

$$D(G) = \max_i(\max_j(\text{distance}(N_i, N_j)))$$

Where:

- $D(G)$ is the diameter
- $\text{distance}(N_i, N_j)$ represents the shortest path between nodes N_i and N_j.

For example, the diameter of the graph shown in Figure 1.10 is 2.

1.3.2.2 Density

Density measures how connected a network is. It is calculated as the ratio of actual edges to the maximum possible edges. The value ranges from 0 (disconnected graph) to 1 (fully connected graph). It helps assess how closely individuals or entities are connected within a network.[1][2][17]

$$D = \frac{|E|}{\frac{|V|(|V|-1)}{2}}$$

where:

- $|E|$: the number of edges in the graph.
- $|V|$: the number of vertices in the graph.

For example, the density of the graph shown in **Figure 1.10** is 0.666666666667.

1.3.2.3 Centrality

Centrality is a measure of the importance or influence of nodes within a network, helping to identify the most impactful elements in information flow and communication. There are several types of centrality, including:

1. **Degree centrality**: It is based on the number of edges (links) a node has with other nodes in the network. The more edges a node has, the greater its importance or influence within the network.[1][2] [18]

$$CD(i) = \frac{d(i)}{n-1}$$

where:

- $CD(i)$: Degree Centrality.

- d(i): The number of edges connected to the node.
- n: the total number of nodes in the graph.

2. **Betweenness centrality:** It measures how much a node controls the flow of information across the network. In other words, it is used to determine how often the shortest paths between any two nodes pass through the target node. The more frequently a node appears on the shortest paths between other nodes, the more important it is as a hub for routing or exchanging information within the network.[1][2] [18]

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st} is the total number of shortest paths from node s to node t.
- $\sigma_{st}(v)$ is the number of those paths that pass through node v.

3. **Closeness centrality**: It reflects how easily a node can reach other nodes in the network. In other words, the smaller the total distance between the node and the rest of the nodes, the higher its closeness centrality. This means the node is more centrally located and has a greater ability to interact quickly with the rest of the network.[1][2] [18]

$$C_c(v) = \frac{1}{\sum_{u \epsilon V \setminus \{V\}} d_G(u,v)}$$

Where: dG(u, v) is the shortest path distance between nodes u and v.

Centrality analysis is used in social network studies to understand influence dynamics and identify key actors within the network. This contributes to analyzing network structures and making decisions based on relationships between different components.

1.3.2.4 Clustering coefficient

The clustering coefficient is a measure of the tendency of nodes within a network to form interconnected groups, reflecting the degree of connectivity among a node's neighbors. High clustering coefficient values indicate the presence of tightly-knit subcommunities within the network, suggesting that individuals or entities tend to form groups with strong internal connections.

This metric is used in social network analysis to understand the structure of small communities within the larger network and to assess the cohesiveness of relationships among its components. It helps in studying the dynamics of interaction within the network.[1][2]

- **The local clustering coefficient** is calculated using the following formula: [19]

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

Where:

- C_i: Local clustering coefficient of node i
- E_i: Number of actual links between the neighbors of node i
- k_i: Number of neighbors of node i

Alternatively, the number of actual links E_i can be calculated as:

$$E_i = \frac{k_i(k_i - 1)}{2}$$

- **The global clustering coefficient** is defined as the number of closed triplets (or 3 × number of triangles) divided by the total number of triplets (both open and closed).[1]

1.3.3 The role of graphs in analyzing social networks

Graphs play a fundamental role in analyzing social networks, as they allow the representation of relationships between individuals in the form of nodes and edges, helping to understand the network's structure and the dynamics of interaction among its components. This representation relies on a set of metrics and analyses that reveal the influence of individuals within the network, such as centrality measures that identify the most influential nodes and their role in facilitating the flow of information. Graphs also contribute to the study of information diffusion across the network, helping to understand how news and ideas spread and identifying the factors influencing their speed and extent of dissemination. Additionally, graphs are used to predict future links, allowing for the analysis of potential trends within the network based on previous interaction patterns. Thanks to these tools, graph analysis provides an accurate insight into the structure of social networks, making it essential for studying social interactions, evaluating influence, and exploring the dynamic patterns that govern the network.[1][20]

1.3.4 Applications of social networks analysis

Social Network Analysis (SNA) is used in a wide range of practical applications, as it helps to understand social interactions, dynamics of influence, and the spread of information within communities. This allows for a deeper understanding of relationships between individuals and institutions and their impact on various decisions and behaviors. These applications span several fields, including:[1]

1. **Political science:**

   Social Network Analysis is a powerful tool for understanding relationships between countries and international organizations, as it can clarify dynamics of power and political influence. It is also used to study networks of political influence and lobbying groups, helping to analyze how political ideas spread and identify key actors in political processes and decision-making.[1]

2. **Economics and business management:**

   Social Network Analysis plays a key role in studying business networks and collaboration between companies. This analysis helps understand the flow of information and resources between organizations, allowing businesses to identify potential commercial opportunities and improve market strategies. It is also used to understand the impact of social relationships on economic performance and business decision-making, especially in environments dependent on relationships and personal networks.[1]

3. **Psychology:**

   Social Network Analysis helps study the impact of social relationships on mental health and individual behavior. By analyzing social support networks, researchers can understand how personal relationships influence stress and depression levels, as well as how habits and psychological tendencies are shaped within individuals' social networks.[1]

4. **Public health:**

   Social Network Analysis is used to study the spread of diseases and infections within communities. By understanding communication patterns between individuals, researchers can design effective strategies to limit the spread of epidemics and promote health awareness. The analysis also helps identify social factors that affect the adoption of healthy behaviors, such as exercising or quitting smoking, by understanding how social influence spreads among individuals.[1]

5. **Education:**

   Social Network Analysis is used to study interactions between students and teachers, helping to understand how social relationships affect academic achievement. These analyses can reveal ways to improve the educational environment and enhance collaboration among students. It is also used to study informal learning networks that contribute to the exchange of knowledge among individuals.[1]

In general, Social Network Analysis is a versatile tool applied across various fields to understand human interactions and enhance effectiveness in decision-making and strategic planning processes.[1]

1.4 Conclusion

In conclusion, social networks have become integral to modern life, influencing personal relationships, business practices, and societal structures. Their evolution from simple platforms for communication to powerful tools for marketing, political influence, and social change demonstrates their far-reaching impact. Through the application of graph theory and social network analysis, we can better understand the complex dynamics of these networks and the roles individuals and groups play within them. As social networks continue to evolve, they will undoubtedly shape the future of how we connect, share information, and navigate the digital world. The continued study of these networks, both from a theoretical and practical perspective, is essential for understanding their influence and harnessing their potential in a rapidly changing global landscape.

Chapter 2: Overlapping Community Detection

2.1 Introduction

In recent years, the analysis of complex networks particularly social networks has gained increasing importance due to the rising availability of data and the need to understand human behavior and interactions. One of the fundamental aspects of this analysis is community detection, which refers to identifying groups of nodes that are more densely connected with each other than with the rest of the network. Traditional methods have primarily focused on non-overlapping communities, assuming that each node belongs to only one group. However, in real-world scenarios, individuals often belong to multiple social circles simultaneously, leading to the emergence of overlapping communities. Detecting these overlapping structures is essential for accurately representing the multi-faceted nature of relationships, enabling deeper insights into information flow, social influence, and network dynamics. This chapter explores the concept, importance, types, algorithms, and challenges of overlapping community detection, highlighting its relevance in various real-world applications and scientific domains.

2.2 Communities

Communities are dense clusters of nodes in a network that are distinguished from other nodes in the same network by shared attributes. Strong connections are frequently seen between nodes that have similar characteristics. This aids in locating groups of closely related nodes that have particular characteristics… [21][22]

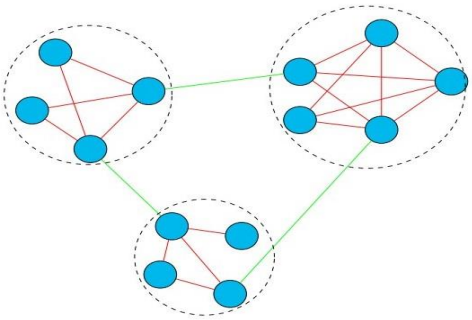

Figure 2.1: A simple graph with three communities [17]

2.2.1 Community detection in social networks

Community detection is the process of identifying interconnected groups within a network, where internal links are stronger than links between different groups. One of the most well-known methods is the Louvain algorithm. Advanced analytical techniques are also used to improve the accuracy and effectiveness of detecting communities, especially in large-scale social networks. [21][22]

2.2.1.1 Definition

Community detection in social networks is the process of dividing the network into groups of nodes with strong internal connections. The goal is to understand the social structure and frequent interactions by identifying groups that interact more among themselves than with the rest of the network.[21][22]

2.2.1.2 Significance

Community detection in social networks is a key tool for understanding how individuals interact within a network. By identifying interconnected groups with strong interactions, we gain a deeper insight into both individual and collective behavior. This detection also improves social data analysis, contributing to better recommendation strategies, such as targeted ads or products, and helps in analyzing social influences, like measuring the impact of individuals on others within the network.[21][22]

2.2.2 Communities structure

The community structure in a network consists of a set of discovered communities, represented as:

$$C = C_1, C_2, C_3, \ldots, C_k$$

Where C represents the overall community structure, and $C_1, C_2, \ldots, C_k$ represent the individual communities that make up this structure.[17]

For example, in Figure 2.2, the communities are as follows:

- $C_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $C_2 = \{10, 11, 12, 13, 14, 15\}$
- $C_3 = \{16, 17, 18\}$

Thus, the community structure is: $C = \{C_1, C_2, C_3\}$.

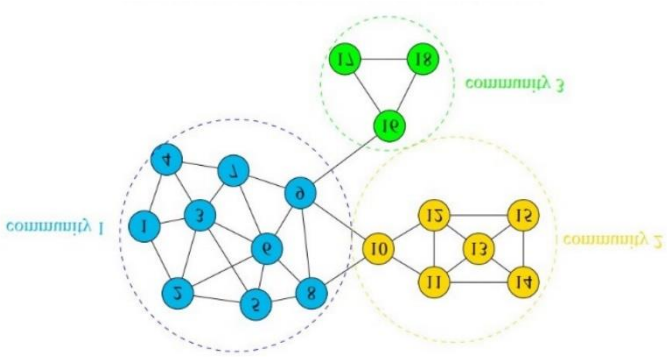

Figure 2.2: The community structure of an example network [17]

And this structure can take two main forms difference between them as shown in the following table:

Feature	**Overlapping Communities**	**Non-Overlapping Communities**
Node Membership	A node may concurrently be a member of multiple communities.	Every node is a member of a single community.
Representation of Social Interactions	Reflects multifaceted interactions and different real-life affiliations.	Simplifies interactions while decreasing the accuracy of social complexity representation.
Modeling Accuracy	More efficient and realistic in simulating contemporary social networks.	Offers a more conventional, easier model.
Ease of Application	More difficult to comprehend and apply.	Simpler to use and evaluate.

Table 2.1: Difference between Overlapping and Non-overlapping Communities [23]

2.2.3 Traditional community detection algorithms

1. **Girvan-newman algorithm:** is a hierarchical, divided method for identifying communities in complex networks is simpler to examine. Edges having the highest edge betweenness centrality that is, edges that frequently occur on the shortest paths between nodes are removed iteratively. The method progressively divides the network into discrete communities by eliminating these "bridge" links, exposing clusters of nodes that

are sparsely connected to other clusters and apply but strongly connected within them selves. [24]

Steps

- Based on removing edges that connect different communities.
- Compute the betweenness centrality for all edges.
- Remove the edge with the highest centrality.
- Recalculate betweenness for remaining edges.
- Repeat the process until the network is divided into disconnected components (communities). [24]

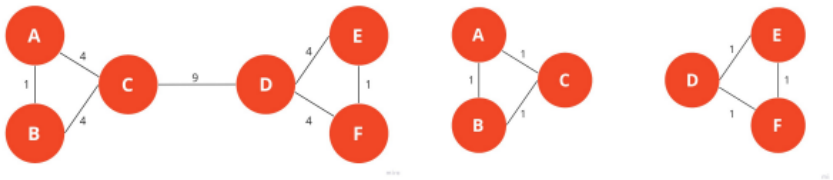

Figure 2.3: Exemple of Girvan-newman algorithm [17]

2. **Louvain method:**

is a greedy optimization technique was created to find non-overlapping communities in big networks fast. It functions by optimizing modularity, which is a metric that compares the number of links inside communities to those across communities. [25]

Steps

- Initially, each node is assigned to its own community.
- Iteratively move nodes to neighboring communities if modularity increases.
- Merge nodes in the same community into super-nodes, forming a reduced graph.
- Repeat the process until modularity can no longer be improved.[25]

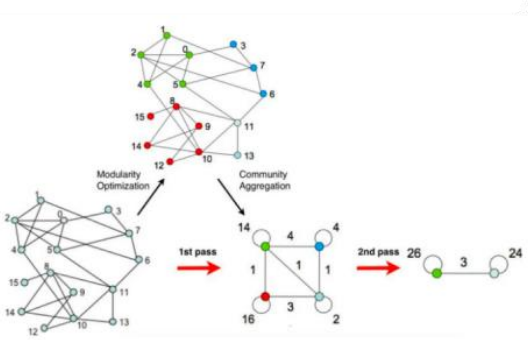

Figure 2.4: Example of two phases in louvain algorithm [17]

And this table presents the advantages and disadvantages of non-overlapping community detection methods.

Method	Advantages	Disadvantages
Girvan-Newman Algorithm	• Effective at detecting distinct, non-overlapping communities. • Produces a clear, hierarchical community structure.	• Computationally expensive for large networks. • Detects only non-overlapping communities (each node in one group).
Louvain Method	• Highly efficient and scalable for large networks. • Tends to find communities with strong internal connections.	• Detects only non-overlapping communities. • May merge small communities due to the resolution limit. • Can produce weakly connected or internally fragmented communities.

Table 2.2: Advantages and disadvantages of non-overlapping community detection methods [24][25][26]

2.2.4 Overlapping community

Overlapping communities represent individuals who belong to multiple groups simultaneously, such as academic, hobby-based, or cultural communities. This reflects the complex nature of social relationships. Detecting these communities provides a deeper understanding of network structure and helps analyze information diffusion and social influence more accurately.[2][27][28]

2.2.4.1 Definition

A collection of communities that have certain nodes in common with one another is referred to as overlapping communities. Based on specific attributes, the nodes are part of one or more communities. Examine node 10, which is a member of both communities c1 and c2, as depicted in Figure. After that, communities c1 and c2 are helped to overlap. [29]

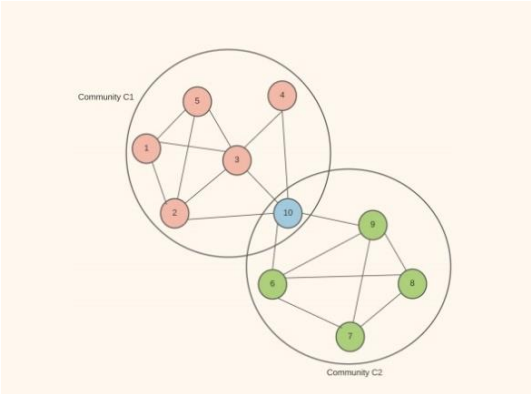

Figure 2.5: Overlapping Communities C1 and C2 sharing a node 10 [29]

2.2.4.2 Significance

Overlapping communities are a crucial aspect of understanding complex social networks, as they reflect the interactive and dynamic nature of human relationships. These communities provide deeper insights into how individuals are connected and how information is exchanged across the various groups they belong to. [2][27]

- **Real-world representation:** Overlapping communities represent real-life situations where individuals or entities are part of multiple groups simultaneously. For example, a researcher who belongs to several research teams, or a person who is a member of both friend and family groups. This reflects the multidimensional nature of social relationships.[28][30]

- **Network dynamics:** In dynamic networks, overlapping communities provide important insights into the evolution of relationships over time, which is crucial for understanding the changes in biological networks, social media networks, and other complex systems.[28]
- **Bridging roles:** Overlapping nodes often serve as "bridges" between different communities, facilitating communication between groups and contributing to the exchange of information and mutual influence.[31]
- **Practical applications:** Overlapping communities play a significant role in various fields, such as public opinion analysis, brain neural network analysis, precision marketing, and drug discovery, due to their complex and interconnected nature.[31]

2.2.5 Benefits of overlapping communities

Overlapping communities provide a rich and dynamic environment that encourages interaction and knowledge exchange among diverse groups, which directly contributes to stronger social ties and collective innovation. One of the most prominent advantages is knowledge sharing, as these interconnected settings offer members access to a broader range of ideas, experiences, and expertise fostering both personal and professional growth. According to Selena S. Blankenship and Wendy E. A. Ruona from the University of Georgia, knowledge sharing within communities leads to improved learning by exposing participants to a wider reservoir of insights and perspectives. [32]

In addition, overlapping communities help build stronger social bonds, as the act of sharing knowledge nurtures reciprocity and trust, thereby enhancing members' sense of belonging and mutual support. [32]

Furthermore, they play a key role in community development and cultural preservation. Through the exchange of traditions, practices, and collective problem-solving, knowledge sharing helps address common issues and maintain shared values across overlapping groups . [33]

2.2.6 Types of overlapping communities

Communities that overlap happen when people or nodes are a part of more than one group at the same time. Here are some typical types:

- **Social circles:** These communities stand in for interpersonal bonds like those with family, friends, coworkers, or hobby groups. People frequently fall into more than one social circle, which causes overlaps between them. [34]

- **Interest-based groups:** Interest-Based Groups: Communities built around common passions, like music lovers, sports teams, or book groups. Members usually belong to more than one interest-based group, which causes memberships to overlap. [35]
- **Professional networks:** When people engage with coworkers from various departments or industries, overlapping communities are created in the workplace. These overlaps are frequently fostered by networking events and cross-functional collaborations.[36][37]

2.2.7 Examples of overlapping communities

The following are examples of overlapping communities in the real world:

- **Online forums:** People frequently take part in several forums pertaining to various subjects, which leads to overlaps across communities with a range of interests. Reddit users, for instance, usually participate in multiple subreddits at the same time. [38].
- **Community organizations:** Volunteers frequently collaborate with a variety of organizations, including neighborhood associations, charitable organizations, and environmental groups. Their participation produces overlapping networks of cooperation and common objectives. [39]
- **Collaborative networks:** Researchers are part of overlapping collaborations across disciplines or projects in academic or professional contexts (such as ArXiv citation networks). [36]

2.2.8 Analyzing overlapping communities

Analyzing overlapping communities is essential for understanding complex networks, particularly when nodes belong to more than one community. Through a review of the key methods and algorithms, effective analysis approaches can be categorized as follows:

- **Distributed Neighborhood Threshold Method (DNTM):** This method relies on analyzing the neighborhood distribution of boundary nodes, starting from non-overlapping communities, and uses a threshold value to determine the impact of neighbors on a node's membership in multiple communities. The main steps include creating initial non-overlapping clusters, analyzing neighborhoods to identify potential overlapping nodes, and updating communities based on these findings. [40]

  Application: It is effective in social networks, tested on real-world datasets, and outperforms other algorithms in terms of quality metrics such as Modularity.

- **Membership Degree Propagation Algorithm (MDPA):** This algorithm improves the accuracy and efficiency of detecting overlapping communities compared to traditional

label propagation methods by using membership degree probabilities rather than fixed labels. [41]

Advantages: It performs better in identifying overlapping nodes in both artificial and real-world datasets, while reducing computational complexity.

- **Link Partitioning Around Medoids (LPAM):** This method focuses on clustering links instead of nodes using distance functions, such as commute distance. If a node's neighboring edges belong to different clusters, it is considered to be part of multiple communities. Example: A person who belongs to both the "colleagues" and "soccer players" communities based on different types of relationships. [30]

  Application: Well-suited for small and medium-sized networks, particularly relational data networks.

- **Node importance and adjacency information:** This approach combines adjacency data with node importance metrics to detect overlapping communities. Random Walk algorithms are used as a starting point to analyze the network's structure and its evolution over time. [41]

  Advantages: It provides a high-precision method for tracking the dynamic changes in the network's structure.[41]

2.3 Overlapping community detection

The detection of overlapping communities represents a fundamental step in understanding the complex structure of social networks, as it reflects the reality in which nodes belong to more than one community.

2.3.1 Objectives of overlapping community detection

The main goal of overlapping community detection is to identify nodes that belong to multiple communities and understand the complex network structures in which this occurs. [28] Key objectives include:

- **Identifying overlapping nodes:** Recognizing nodes that belong to several communities, reflecting real-world situations where individuals have multiple roles or affiliations. [28][30]

- **Increasing computational efficiency:** Developing algorithms that efficiently handle complex networks with overlapping communities, improving both speed and accuracy. [28][42]

- **Improving understanding of community structure:** Analyzing the dynamics and evolution of overlapping communities over time to better understand network behavior. [30] [41]
- **Multi-objective optimization:** Using various criteria to evaluate community structures, ensuring algorithms adapt to different network properties and provide comprehensive insights. [42]

2.3.2 Applications of overlapping community detection

Overlapping community detection is widely used across multiple fields to identify nodes that belong to more than one community simultaneously, reflecting the complex nature of real-world networks. Key applications include:

- **Social network analysis:** Helps understand complex social interactions and how communities evolve over time. [41]
- **Biological systems:** Identifies biological elements (e.g., proteins) involved in multiple functional pathways.[28]
- **Academic collaboration networks:** Analyzes researcher collaborations across various disciplines. [43]
- **Information networks:** Improves content recommendation by identifying documents or websites with multiple thematic affiliations. [43]
- **Marketing and recommendation systems:** Enhances targeted marketing by addressing users with diverse interests. [30]
- **Network dynamics:** Supports the study of how networks change and adapt over time. [41]
- **Big data environments:** Applies scalable algorithms like BIGCLAM and LPAM to large-scale networks.[30]
- **Machine learning applications:** Aids in tasks like node classification and link prediction using graph neural networks. [44]

2.3.3 Classification overlapping community detection algorithm

Overlapping Community Detection algorithms are classified based on the fundamental methods and strategies they rely on. Below is the main classification of these algorithms:

Category	Description	Examples
Node-Based Algorithms	Focus on clustering nodes based on their structural properties.	- Clique Percolation Method (CPM) - DEMON
Link-Based Algorithms	Focus on clustering links rather than nodes to detect communities.	- Link Partitioning - NDOCD
Seed-Based Methods	Start with "seed" nodes and grow communities around them.	- Distributed Neighborhood Threshold Method (DNTM)
Ensemble-Based Methods	Create overlapping communities by combining results from disjoint algorithms.	- MEDOC - EnCoD
Multi-Objective Optimization	Use optimization techniques to simultaneously optimize multiple objectives in community detection.	- MOEA-OCD - MCMOEA - MOOCD-RC

Table 2.3: Classification overlapping community detection algorithm [28][45][46]

2.3.4 Challenges in overlapping community detection

Detecting overlapping communities in social networks presents several complex challenges that affect the accuracy and efficiency of the results. The main challenges include:

- **Computational complexity** is a major issue. Analyzing large and dense networks requires intensive computation, leading to slow execution and high resource consumption. To address this, researchers aim to develop more efficient algorithms that reduce reliance on external parameters. [28][30][46]
- **Shadowing effect** occurs when larger or denser communities obscure the presence of smaller ones, making them harder to detect. A possible solution is to use cascade detection methods that iteratively remove the larger communities, allowing the smaller ones to emerge more clearly. [46]

- **Ambiguity in the definition of a community:** The concept of "community" varies across algorithms and application contexts, resulting in inconsistent outcomes and making it difficult to compare studies. To overcome this standardizing terminology or adopting flexible frameworks that support multiple interpretations is essential. [46]
- **Scalability** remains a significant challenge. Handling large-scale networks efficiently is difficult, often causing delays or incomplete results. This can be addressed through parallel computing and the development of scalable algorithms. [46]
- **Overlap density and diversity** also complicate detection. High levels of node sharing and varying patterns of overlap make it difficult to define clear community boundaries. Advanced algorithms that explicitly account for overlap properties are needed to tackle this problem. [46]
- Many methods suffer from a **dependence on prior information** relying on predefined parameters or assumptions about community structure can lead to biased results and limit generalizability. Therefore, data-driven approaches that minimize the need for prior assumptions are recommended. [46]

2.4 Conclusion

Overlapping community detection is a crucial advancement in network science, bridging the gap between theoretical models and the complexity of real-world interactions. By recognizing that nodes especially in social networks often participate in multiple communities, researchers and practitioners can achieve a more accurate and nuanced understanding of network behavior. This chapter has discussed the foundations of community structures, reviewed traditional and modern detection algorithms, examined the various types and benefits of overlapping communities, and highlighted practical applications in diverse fields such as biology, marketing, and information systems. Despite its progress, overlapping community detection continues to face significant challenges, including computational complexity, information overload, and scalability issues. Addressing these challenges through advanced algorithms and optimization strategies will be key to unlocking the full potential of this vital area of study in the evolving landscape of complex networks.

Chapter 3: Comparison and Results

3.1 Introduction

This chapter aims to present an experimental study to evaluate the performance of overlapping community detection algorithms in social networks. After a detailed review of the theoretical aspects in the previous chapters, this section focuses on applying the selected algorithms to real-world and benchmark networks, followed by analyzing and comparing the results using appropriate evaluation metrics. The goal of this analysis is to highlight the strengths and weaknesses of each algorithm in various network contexts.

3.2 Evaluation metrics

In order to assess the performance of overlapping community detection algorithms, several evaluation metrics are used.

3.2.1 Modularity(Q)

Modularity is a widely used metric for evaluating the quality of community structures in networks. It measures the strength of division of a network into communities by comparing the density of edges inside communities to the density expected in a random network with the same degree distribution. A higher modularity value indicates a stronger community structure.[17]

The modularity score Q is defined by the following formula:

$$Q = \frac{1}{2m}\sum(A_{ij} - \frac{k_i.k_j}{2m})\delta(c_i.c_j)$$

Where:

- A_{ij}: Represents an element of the network's adjacency matrix. It takes the value 1 if there is a link between nodes i and j, and 0 otherwise.
- k_i and k_j: Represent the degrees of nodes i and j, respectively (i.e., the number of edges connected to each node).
- m: Denotes the total number of edges in the network.
- c_i and c_j: Indicate the communities to which nodes i and j belong, respectively.

- $\delta(c_i, c_j)$: Is the Kronecker delta function, defined as follows:

$$\delta(c_i . c_j) \begin{cases} 1, & if\ nodes\ i\ and\ j\ belong\ to\ the\ same\ community \\ 0, & Otherwise \end{cases}$$

3.2.1.1 Extended modularity for overlapping communities

Extended Modularity is an adaptation of the traditional modularity metric designed to evaluate overlapping community structures in social networks. as showen in Equation 3.1 Unlike standard modularity, it takes into account the fact that nodes may belong to multiple communities. It measures how much the density of links inside communities exceeds what would be expected in a random network, considering shared node memberships. [47]

- The extended modularity is defined as follows: [47]

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] . \frac{1}{T_i T_i} \sum_{c \epsilon C_i \cap C_j} 1 \qquad (3.1)$$

where:

- A_{ij} is the element of the adjacency matrix between nodes i and j,
- k_i and k_j are the degrees of nodes i and j,
- m is the total number of edges in the network,
- C_i and C_j are the sets of communities that nodes i and j belong to, respectively,
- T_i and T_j are the numbers of communities node i and node j belong to.

3.2.2 Normalized mutual information (NMI))

NMI is a standard metric used to evaluate the similarity between two partitions of a network. It is widely applied in community detection to compare the detected partition with a ground-truth partition. NMI is based on information theory and measures how much information is shared between two partitions, with values ranging from 0 (no similarity) to 1 (perfect match). [17]

For two partitions X and Y of a network, the value of NMI is calculated using the following equation:

$$NMI(X,Y)\ \frac{\sum_{i=1}^{C_x}\sum_{j=1}^{C_Y} N_{ij}log(\frac{N_{ij}N}{N_iN_j})}{\sum_{i=1}^{C_x} N_ilog(\frac{N_i}{N}) + \sum_{j=1}^{C_Y} N_jlog(\frac{N_j}{N})}$$

Where:

- X: Represents the actual (ground-truth) partition of the network.
- Y : Denotes the partition discovered by the community detection algorithm.
- C_X: Is the number of communities in partition X.
- C_Y : Is the number of communities in partition Y .
- N: Represents the total number of vertices in the network.
- N_{ij}: Is the number of vertices shared between community i in partition X and community j in partition Y.
- N_i: Is the number of vertices in community i of partition X (i.e., the sum of row i in the N_{ij} matrix).
- N_j: Is the number of vertices in community j of partition Y (i.e., the sum of column j).

3.2.2.1 Overlapping normalized mutual information (ONMI)

ONMI is an extension of the traditional NMI metric that is specifically designed to handle overlapping community structures in networks, where nodes may belong to multiple communities. This metric accurately reflects the degree of overlap and serves as a standard tool for assessing the performance of overlapping community detection algorithms. [48]

For two overlapping partitions X and Y of a network, the value of ONMI is calculated by the following equation: [48]

$$ONMI(X,Y) = \frac{2 \cdot I(X;Y)}{H(X) + H(Y)} \quad (3.2)$$

where:

- X represents the ground-truth overlapping partition of the network,
- Y represents the overlapping partition detected by the community detection algorithms,
- $I(X;Y)$ is the mutual information between partitions X and Y , adapted for overlap-

ping communities.

- H(X) and H(Y) are the entropies of partitions X and Y, respectively, considering overlapping memberships.

The ONMI metric ranges from 0 to 1. A value closer to 1 indicates a high similarity between the overlapping partitions, while a value near 0 reflects a low similarity. Specifically, if the partitions X and Y are identical, then $ONMI(X, Y) = 1$; whereas if they are completely different, then $ONMI(X, Y) = 0$. [48]

3.3 Overlapping community detection algorithms

Among the algorithms used for detecting overlapping communities in social networks:

3.3.1 SLPA (Speaker-Listener Label Propagation Algorithm)

SLPA is a label propagation algorithm based on a speaker-listener interaction model. Each node can belong to multiple communities depending on the frequency of received labels. [49]

Steps:

1. Initialize each node with a unique label.
2. For a number of iterations:
    - Each listener selects neighbors as speakers.
    - Each speaker sends a label probabilistically based on its memory.
    - The listener accepts and stores one of the received labels.
3. After all iterations, compute the label distribution for each node.
4. Apply a threshold to retain the most frequent labels and determine overlapping communities. [49]

3.3.2 CPM (Clique Percolation Method)

CPM (Clique Percolation Method) is a technique for detecting overlapping communities in networks by identifying fully connected groups of nodes called *k*-cliques. A community is defined as a group of *k*-cliques that are connected through shared *k--1* nodes, allowing nodes to belong to multiple communities. [50]

Steps:

1. **Extract all *k*-cliques:** Identify all fully connected subgroups (cliques) of size *k* in the network.
2. **Build a clique graph:** Represent each *k*-clique as a node in a new graph.
3. **Connect adjacent cliques:** Add an edge between two *k*-cliques if they share *k--1* nodes.
4. **Detect communities:** Each connected component in this new graph represents a community in the original network. [50]

Algorithm 1 CPMZ pseudocode

```
UF → Empty Union-Find data structure
Setz → Empty Dictionary
for each k-clique c_k ∈ G do
    S → ∅                                   ▷ Sets of z-cliques to merge
    for each (k − 1)-clique c_{k−1} ⊂ c_k do
        P → ∅
        for each z-clique c_z ⊂ c_{k−1} do
            Setz[c_z] → {UF.Find(p) | p ∈ Setz[c_z]}
            if P == ∅ then
                P → Setz[c_z]
            else
                P → P ∩ Setz[c_z]
            end if
        end for
        S → S ∪ P
```

```
16:     end for
17:     q → NULL                          ▷ Identifier of the resulting set of z-cliques
18:     if S == ∅ then
19:         q → UF.MakeSet()
20:     else
21:         q → UF.Union(S)
22:     end if
23:     for each z-clique c_z ⊂ c_k do
24:         Setz[c_z].add(q)
25:     end for
26: end for
```

[50]

3.3.3 BigCLAM (Cluster Affiliation Model for Big Networks)

BigCLAM is a model for detecting overlapping communities in large networks. It represents each node's affiliation to different communities with non-negative values indicating the strength of membership. The model assumes that the probability of an edge between two nodes increases with the number of shared communities, and these membership values are learned by optimizing a likelihood function that best represents the network. [43]

Steps:

1. **Initialize membership values:** Assign initial non-negative values to each node for each community, representing the node's strength of membership.
2. **Optimize membership values:** Use an optimization algorithm (e.g., gradient ascent) to adjust the membership values to better represent the network.
3. **Calculate edge probabilities:** Model the probability of an edge between nodes u and v using the formula

   $$p(u, v) = 1 - \exp(-F_u \cdot F_v^T)$$

   where F_u and F_v are the membership vectors of nodes u and v.
4. **Iterate optimization:** Repeat the optimization steps until convergence is reached and the communities are well represented. [43]

3.3.4 DEMON (Democratic Estimate of the Modular Organization of a Network)

DEMON is an algorithm for detecting overlapping communities in complex networks. It follows a local-first approach, analyzing the network from the perspective of each node individually. It was developed by Michele Coscia et al. and aims to identify overlapping communities by focusing on local network structures. [51]

Steps:

1. Extract the ego network of each node (its neighbors only).
2. Apply label propagation on each ego network to find local communities.
3. Merge similar local communities based on a similarity threshold.
4. Aggregate the merged communities to form the global overlapping communities in the network. [51]

Algorithm 2 The pseudo-code of DEMON algorithm.

Require: G : (V, E); $C = \emptyset$; $\epsilon \in [0..1]$
Ensure: set of overlapping communities C

```
for all v ∈ V do
    e → EgoMinusEgo(v, G)
    C(v) → LabelPropagation(e)
    for all C ∈ C(v) do
        C → C ∪ v
        C → Merge(C, C, ϵ)
    end for
end for
return C
```

[51]

3.3.5 Ego-Splitter(Overlapping Community Detection via Edge Label Propagation)

Ego-Splitter is an algorithm for detecting overlapping communities in complex networks. It uses a local-first approach by analyzing the ego-net of each node and creating multiple (persona) nodes that represent the different roles or communities the node belongs to, thus allowing for overlapping memberships. [52]

Steps:

1. Extract the ego-network of each node, which includes the node itself and its immediate neighbors.
2. Split the ego-network into connected components.
3. Assign each connected component to a separate *persona* of the original node.
4. Construct a new network where each persona is treated as an independent node.
5. Apply a standard community detection algorithm (e.g., Label Propagation or Louvain) to detect communities in the persona graph. [52]

3.4 Experimental results and analysis

This section presents the experimental results obtained from applying overlapping community detection algorithms on various networks, followed by a detailed analysis and comparison based on evaluation metrics and execution performance.

3.4.1 Working environment

All experiments were conducted in an environment dedicated to data analysis and testing algorithms on graph-structured data. A platform was used that provides powerful tools for data processing, code execution, and package installation. The working environment is divided into two main aspects: Hardware and Software, which will be detailed in the following subsections.

3.4.1.1 Hardware environment

The experiments were conducted on the Kaggle platform. Hardware specifications included.

- **Session Type:** Draft Session (no hardware accelerator)
- **Maximum Session Duration:** 12 hours
- **RAM:** 30 GiB available (595.1 MiB used during the experiment)
- **Disk:** 57.6 GiB available (2.5 GiB used)

This setup was sufficient to run Python-based graph algorithms efficiently.

3.4.1.2 Software environment

The software environment was entirely based on the Kaggle Kernel environment, which supports execution of Python code in Jupyter Notebooks. Key features of the environment include:

- **Platform:** Kaggle (cloud-based execution)
- **Programming language:** Python 3
- **Notebook environment:** Jupyter-based interface with integrated file system, output viewer, and execution monitoring

This environment enabled seamless development and testing of community detection algorithms without the need for local installations.

Libraries using:

Library	Description
cdlib	Community detection library providing implementations for CPM, DEMON, SLPA, Ego-Splitter, and evaluation metrics like ONMI and Extended Modularity. https://cdlib.readthedocs.io/en/latest/
networkx	Toolkit for creating and manipulating complex networks. Used for graph loading, edge management, and basic network analysis. https://networkx.org/documentation/stable/index.html
numpy	Library for numerical computations and matrix operations. Essential for linear algebra in algorithms like BigClam. https://www.w3schools.com/python/numpy/numpy_intro.asp
pandas	Data manipulation and analysis library. Used for reading datasets (e.g., .csv, .dat) and handling tabular data efficiently. https://pandas.pydata.org/docs/
igraph	High-performance graph library (especially in conjunction with cdlib) used for scalable community detection and graph analysis. https://python.igraph.org/en/stable/

matplotlib	Visualization library used to plot network graphs and statistical charts, helping in understanding algorithm outputs. https://matplotlib.org/stable/index.html
karateclub	A library for unsupervised learning on graph-structured data. Used particularly for running the BigClam algorithm. https://karateclub.readthedocs.io/en/latest/
collections	Built-in Python module providing specialized data structures like defaultdict and Counter to manage node-community mappings and frequency counts. https://docs.python.org/3/library/collections.html
time	Standard Python module used for measuring execution time of algorithms and performance profiling. https://www.programiz.com/python-programming/time
subprocess	Module used to run external processes or scripts, useful when integrating non-python tools or calling shell commands during experiments. https://realpython.com/python-subprocess/

Table 3.1: The libraries used in our work

3.4.2 Datasets

As part of this work, we conducted a comparative study of five algorithms designed for overlapping community detection in social networks: SLPA, CPM, BigCLAM, DEMON, and Ego-Splitter. This study was carried out on eight networks, including five real-world networks (Karate Club, Dolphins, Political Books, Misérables, and Facebook) and three synthetic networks.

To evaluate the performance of these algorithms, we relied on two widely used metrics in this field: Extended Modularity and Overlapping Normalized Mutual Information (ONMI), in addition to measuring the execution time of each algorithm. This study aims to analyze and identify the strengths and weaknesses of each algorithm individually, with a focus on their efficiency in detecting overlapping community structures within networks.

3.4.2.1 Real world dataset

For real-world networks, we used five datasets. Among them, three networks have non-overlapping ground-truth communities: **Karate Club**, **Dolphins**, and **Polbooks**. One network, **Facebook Ego-Network**, contains overlapping ground-truth communities. Additionally, we used the **Misérables** network, which does not have any predefined ground-truth communities.

- **Karate club network**: This is a small social network of 34 nodes representing members of a karate club and 78 edges representing friendships among them, naturally divided into 2 communities. It is widely used to test community detection algorithms, as it reflects a real split due to a conflict that led to the separation of members. [53]
- **Dolphins network**: This network represents 62 dolphins as nodes and 159 edges representing associations or interactions among them. The network is typically divided into 2 or 3 communities and reflects natural grouping based on social behavior or gender.[54]
- **Polbooks network**: This dataset includes 105 nodes representing American political books and 441 edges representing relationships such as co-purchase or shared interests. The network reflects political orientation-based communities (e.g., liberal, conservative), and is divided into 3 well-known communities. [55]
- **Facebook ego-network**: This social network consists of 333 nodes representing Facebook users and about 2,512 edges representing friendships. It is mainly used to study overlapping communities and social structures in online networks. [56]
- **Misérables network**: Based on Victor Hugo's novel, this network includes 77 nodes representing characters and 254 edges for co-occurrences in the same chapters. It is used for analyzing narrative structures in literature. However, no ground-truth or overlapping communities are provided. [57]

3.4.2.2 Generated dataset

The LFR model, a synthetic data generation tool, provides heterogeneity in the distribution of community sizes and node degrees, which are characteristics commonly observed in real-world networks. Additionally, the model offers a wide range of parameters to control the network topology. Most of the parameters were configured as shown in the table below. Node degrees and community sizes are controlled using power-law distributions, with exponents τ and τ respectively. The data in the table below illustrates the parameters used to gen- erate synthetic LFR networks of different sizes: small (100 nodes), medium (1000 nodes), and large (10,000 nodes), Parameters such as the number of nodes, maximum degree, and number of overlapping nodes increase gradually with the size of the network, allowing for realistic simulation of real-world network properties. Some parameters, like the mixing parameter (mu) and the number of communities per node (om), are kept constant to ensure fair comparison across networks.
These parameters help precisely control the network characteristics to closely resemble real-world networks.[58]

Parameter	Dataset1	Dataset2	Dataset3	Description
N	100	1000	10000	Number of nodes
k	10	15	20	Average degree
maxk	20	50	100	Maximum degree
mu	0.3	0.3	0.3	Mixing parameter
t1	2	2	2	Degree distribution exponent
t2	1.5	1	0.8	Community size distribution exponent
minc	10	40	100	Minimum community size
maxc	20	60	300	Maximum community size
on	10	50	500	Overlapping nodes
om	2	2	2	Communities per node

Table 3.2: Overview of parameter selection for LFR networks

3.4.3 Comparison and results

3.4.3.1 real world dataset

- **Karate club network results**

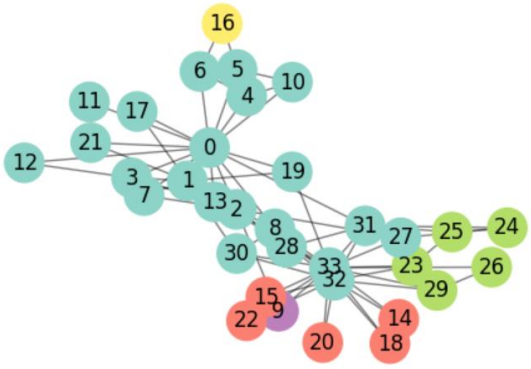

Figure 3.1: Detection of overlapping communities in the Karate network using SLPA (Execution 1)

Figure 3.1 show the overlapping community detection results in Karate network. It actually contains two communities, SLPA algorithm it devided into 5 overlapping communities. With Extended Modularity Q=0.3732 and it's the minimum Q, and ONMI= 0.5805 it's the maximum value.

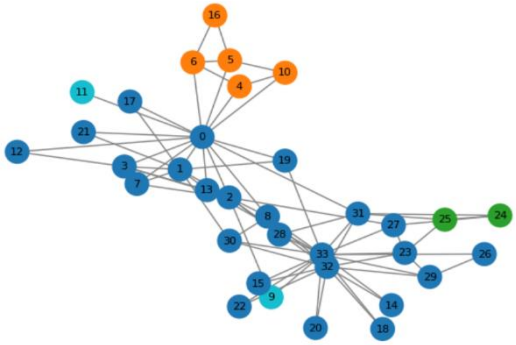

Figure 3.2: Detection of overlapping communities in the Karate network using CPM (Execution 1)

Figure 3.2 show the overlapping community detection results in Karate network. It actually contains two communities, CPM algorithm it devided into 5 overlapping com-

munities. With Extended Modularity Q= 0.9281 and it's the maximum Q, and ONMI= 0.1680 it's the minimum value with DEMON algorithm

- **Facebook network results**

Figure 3.3: Detection of overlapping communities in the Facebook network using Ego-splitter (Execution 1)

Figure 3.3 show the overlapping community detection results in Facebook network. Ego-splitter algorithm it devided into 95 overlapping communities. With Extended Mod ularity Q= 0.0600 and it's the medium Q, and ONMI= 0.2435 it's the maximum value with DEMON algorithm.

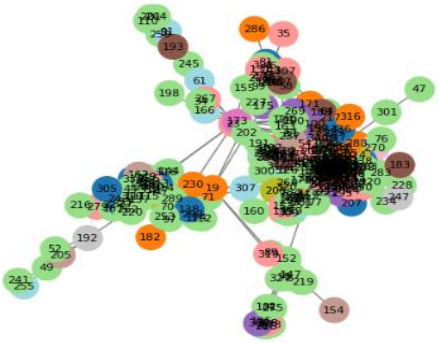

Figure 3.4: Detection of overlapping communities in the Facebook network using SLPA (Execution 1)

Figure 3.4 show the overlapping community detection results in Facebook network. SLPA algorithm it devided into 45 overlapping communities. With Extended Modularity Q= -0.0135 and it's the minimum Q, and ONMI= 0.0167 it's a small value with Bigclam algorithm.

- **Miserables network result**

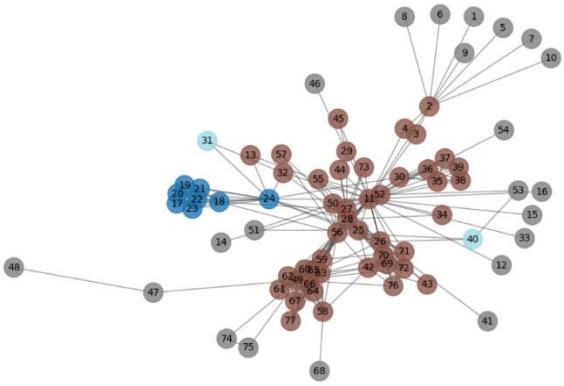

Figure 3.5: Detection of overlapping communities in the Miserables network using DEMON (Execution 1)

Figure 3.5 show the overlapping community detection results in Miserables network. DEMON algorithm it devided into 8 overlapping communities. With Extended Modularity Q= 0.0802 and it's the minimum Q, and ONMI=0.0276 it's the minimum value.

- **Dolphins network results**

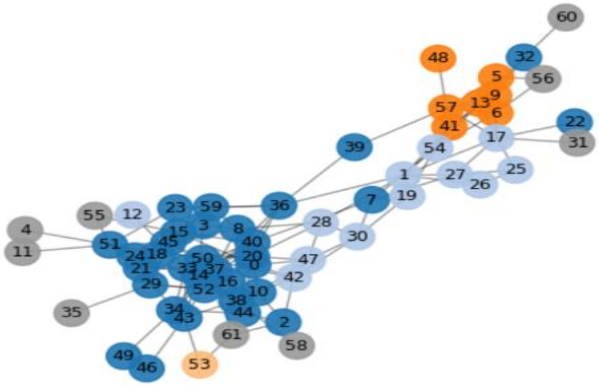

Figure 3.6: Detection of overlapping communities in the Dolphins network using BigClam (Execution 1)

Figure 3.6 show the overlapping community detection results in Dolphins network. BigClam algorithm it devided into 8 overlapping communities. With Extended Modularity Q= 0.2149 and it's the minimum Q, and ONMI= 0.1666 it's the minimum value.

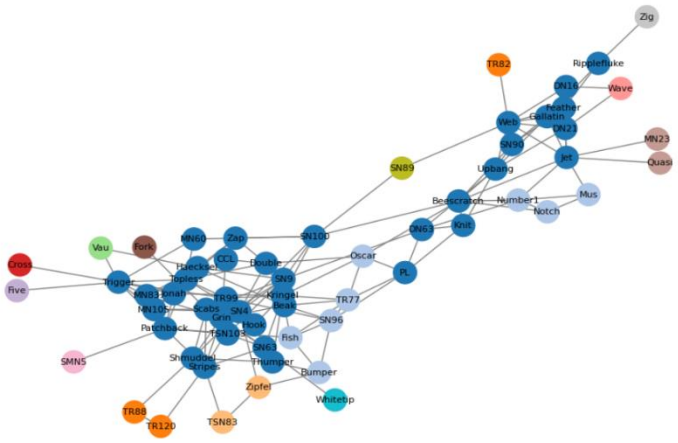

Figure 3.7: Detection of overlapping communities in the Dolphins network using Ego-splitter (Execution 1)

Figure 3.7 show the overlapping community detection results in Dolphins network. Ego-splitter algorithm it devided into 8 overlapping communities. With Extended Modularity Q= 0.7176 and it's the maximum Q, and ONMI= 0.3859 it's a high value.

Comparison:

The performance of five algorithms CPM, SLPA, BigClam, DEMON, and Ego-Splitter was evaluated on five real-world networks: Karate Club, Dolphins, Misérables, Political Books, and Facebook. Three key metrics were used for evaluation: Overlapping Normalized Mutual Information (ONMI), Extended Modularity, and Execution Time.

Overlapping normalized mutual information (ONMI)

The SLPA algorithm achieved the highest ONMI scores on three networks:

- Karate (0.5805)
- Les Misérables (0.9170)
- Political Books (0.4959)

On the Dolphins network, DEMON outperformed others with an ONMI of 0.4539. For the Facebook network, overall ONMI values were lower; however, DEMON led with 0.2898, followed by Ego-Splitter at 0.2435. SLPA showed weaker performance on this network.

Network	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Karate	0.1680	**0.5805**	0.2036	0.1612	0.3485
Dolphins	0.2639	0.3042	0.1666	**0.4539**	0.3859
Misérables	0.8046	**0.9170**	0.0276	0.1244	0.6036
Polbook	0.3338	**0.4959**	0.0952	0.4276	0.4859
Facebook	0.0317	0.0167	0.0073	**0.2898**	0.2435

Table 3.3: ONMI scores for each algorithm on the real-world networks.

Extended modularity

- CPM achieved the highest extended modularity on the Karate network with a score of 0.9281, significantly outperforming the others.
- On the Dolphins network, Ego-Splitter achieved the best extended modularity at 0.7176.
- SLPA performed well on Les Misérables, scoring 0.4933, ahead of the other algorithms.
- On Political Books and Facebook networks, modularity scores were generally lower, with CPM and DEMON showing slight advantages respectively.

Network	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Karate	**0.9281**	0.3732	0.7451	0.7387	0.4421
Dolphins	0.4226	0.4914	0.2149	0.2515	**0.7176**
Misérables	0.1856	**0.4933**	0.0802	0.1530	0.2742
Polbook	**0.4545**	0.4494	0.1498	0.1693	0.2114
Facebook	0.0090	-0.0135	0.0009	**0.1719**	0.0600

Table 3.4: Extended Modularity values for each algorithm on the real-world networks.

Execution time

- CPM was the fastest algorithm across most networks, with execution times ranging from 0.0003 seconds (Karate) to 6.98 seconds (Facebook). Although slower on Facebook, CPM remained relatively efficient compared to others.
- BigClam was generally the slowest, especially on the Political Books and Facebook networks, with execution times up to 0.88 seconds.
- Ego-Splitter and DEMON demonstrated a good balance between accuracy and execution time, providing moderate runtimes with acceptable performance.

- Although SLPA showed superior detection accuracy, it required longer execution times compared to CPM, especially on larger networks.

Network	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Karate	0.0003	0.0152	**0.0817**	0.0200	0.0169
Dolphins	0.0288	0.1460	**0.2807**	0.0300	0.0256
Misérables	0.0915	0.0446	**0.1876**	0.0498	0.0176
Polbook	0.0725	0.0952	**0.2721**	0.1700	0.0295
Facebook	**6.9800**	0.3773	0.8800	0.5832	0.1300

Table 3.5: Execution time (in seconds) for each algorithm on the real-world networks.

Results

- SLPA stands out as the most accurate algorithm in terms of detection quality (ONMI) across most datasets, making it highly suitable when precision is a priority.
- CPM excels in execution speed, especially on smaller networks, making it ideal for time-sensitive applications.
- Ego-Splitter and DEMON offer a balanced trade-off between accuracy and runtime, with Ego-Splitter particularly strong in modularity on the Dolphins network.
- BigClam generally trails in both speed and accuracy but might be appropriate depending on specific application needs.

These findings highlight the importance of selecting the most appropriate algorithm based on network size, accuracy requirements, and computational constraints, The following figures illustrate this :

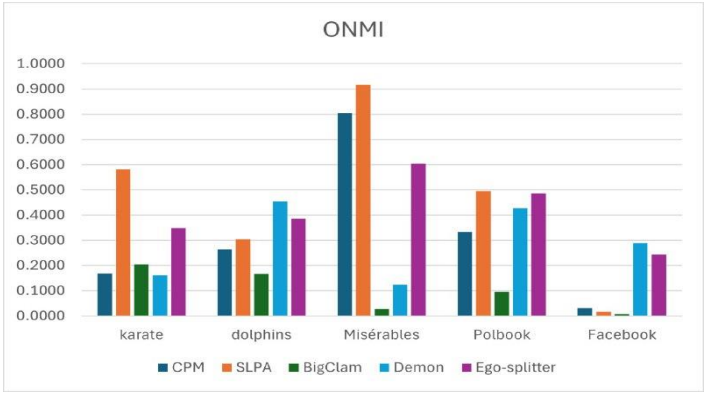

Figure 3.8: Comparison of Overlapping Normalized Mutual Information (ONMI)

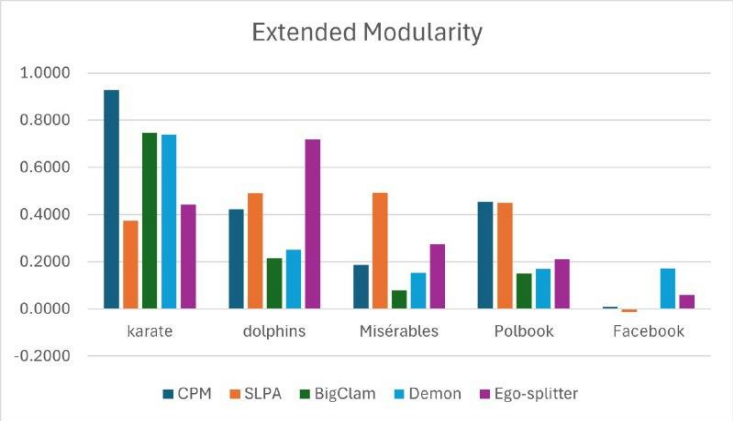

Figure 3.9: Comparison of extended modularity for overlapping communities

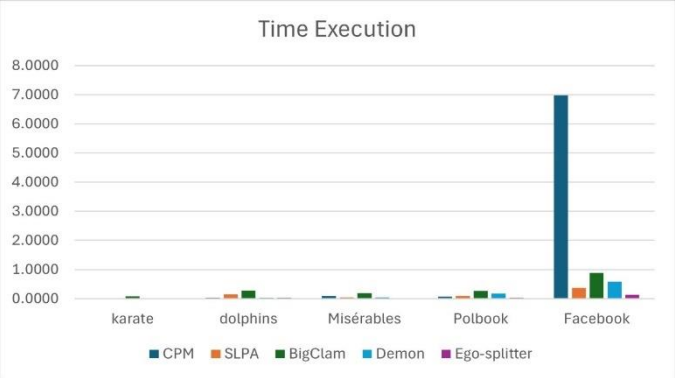

Figure 3.10: Comparison of execution time

3.4.3.2 Generated dataset

- **Dataset1 results**

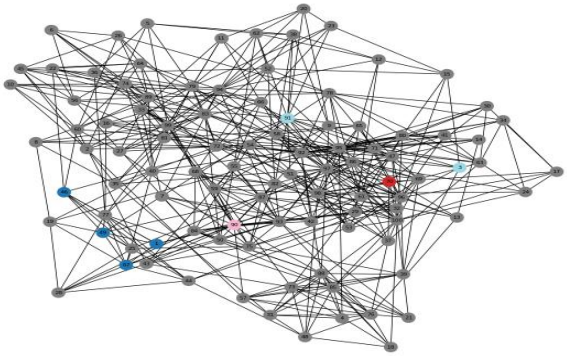

Figure 3.11: Detection of overlapping communities in the Dataset1 using SLPA (Execution 1)

Figure 3.11 show the overlapping community detection results in Dataset1. SLPA algorithm it devided into 8 overlapping communities. With Extended Modularity Q= - 0.3284 and it's a small Q, and ONMI= 0.02669 it's a minimum value.

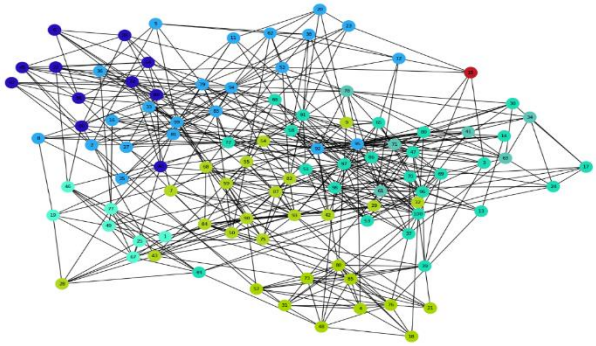

Figure 3.12: Detection of overlapping communities in the Dataset1 using Ego-splitter (Execution 1)

Figure 3.12 show the overlapping community detection results in Dataset1. Ego-splitter algorithm it devided into 94 overlapping communities. With Extended Modularity Q= 0.05121 and it's a small Q, and ONMI= 0.7746 it's a maximum value.

- **Dataset2 results**

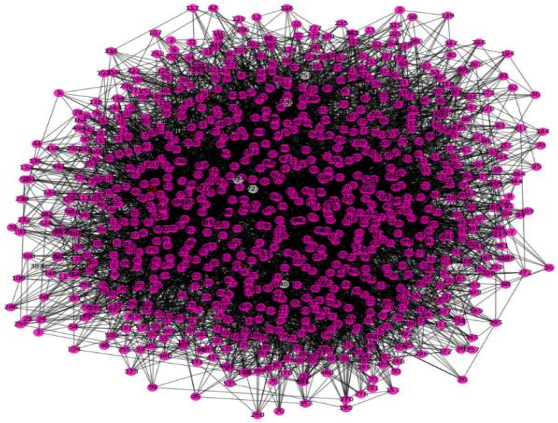

Figure 3.13: Detection of overlapping communities in the Dataset2 using CPM (Execution 1)

Figure 3.13 show the overlapping community detection results in Dataset2. CPM algorithm it devided into 19 overlapping communities. With Extended Modularity Q= -0.0000 and it's a minimum Q, and ONMI= 0.0037 it's a minimum value.

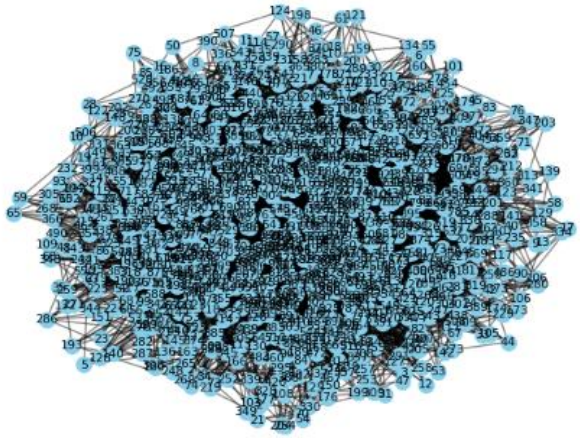

Figure 3.14: Detection of overlapping communities in the Dataset2 using SLPA (Execution 1)

Figure 3.14 show the overlapping community detection results in Dataset2. SLPA algorithm it devided into 22 overlapping communities. With Extended Modularity Q= 0.3113 and it's a high Q, and ONMI= 0.8138 it's a medium value.

Comparison:

Five overlapping community detection algorithms CPM, SLPA, BigCLAM, DEMON, and Ego-Splitter were applied to three synthetic datasets. The evaluation was based on three main metrics: Overlapping Normalized Mutual Information (ONMI), Extended Modularity, and Execution Time.

Overlapping normalized mutual information (ONMI)

DEMON clearly outperformed the others on Dataset1 (0.9919) and Dataset3 (0.8600), while Ego-Splitter achieved the highest accuracy on Dataset2 (0.8898). SLPA also showed competitive results on Dataset2 and Dataset3, although it performed poorly on Dataset1. In contrast, CPM demonstrated consistently weak performance across all datasets, with ONMI values not exceeding 0.03.

Dataset	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Dataset1	0.02669	0.09037	0.6666	**0.9919**	0.7746
Dataset2	0.0037	0.8138	0.0247	0.4520	**0.8898**
Dataset3	0.0179	0.7598	0.1745	**0.8600**	0.8471

Table 3.6: ONMI scores for each algorithm on the generated datasets.

Extended modularity

SLPA achieved the best results on Dataset1 (0.4978) and Dataset3 (0.6364), whereas Ego-Splitter outperformed others on Dataset2 (0.8886). CPM produced negative or near-zero modularity scores, indicating poor alignment with the ground-truth communities in synthetic networks. Additionally, BigCLAM performed poorly in terms of modularity, showing negative scores across all datasets.

Dataset	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Dataset1	-0.3284	**0.4978**	-0.0033	0.0833	0.05121
Dataset2	-0.0000	0.3113	-0.0011	0.3153	**0.8886**
Dataset3	0.0002	**0.6364**	-0.0477	0.0665	0.0134

Table 3.7: Extended Modularity scores for each algorithm on the generated datasets.

Execution time

CPM and Ego-Splitter were the fastest on Dataset1, with CPM taking only 0.0562 seconds. However, execution times increased significantly with larger datasets (Dataset2 and Dataset3). BigCLAM and SLPA were the slowest, especially on Dataset3, where they took 55.26 seconds and 18.82 seconds, respectively. In contrast, Ego-Splitter and DEMON maintained a more balanced performance between accuracy and efficiency.

Dataset	CPM	SLPA	BigClam	DEMON	Ego-Splitter
Dataset1	0.0562	**0.3817**	0.28	0.04	0.04
Dataset2	**9.0896**	5.9229	2.7298	1.0515	0.7039
Dataset3	0.0038	18.82	**55.26**	19.75	14.48

Table 3.8: Execution time (in seconds) for each algorithm on the generated datasets.

Results

DEMON demonstrated superior accuracy across synthetic environments, while SLPA excelled in terms of modularity. Ego-Splitter once again proved to be a balanced choice, offering good detection quality with reasonable execution times. These diverse results highlight

the importance of selecting an algorithm that aligns with the specific characteristics and size of the target network, as well as the desired performance metrics, The following figures illustrate this :

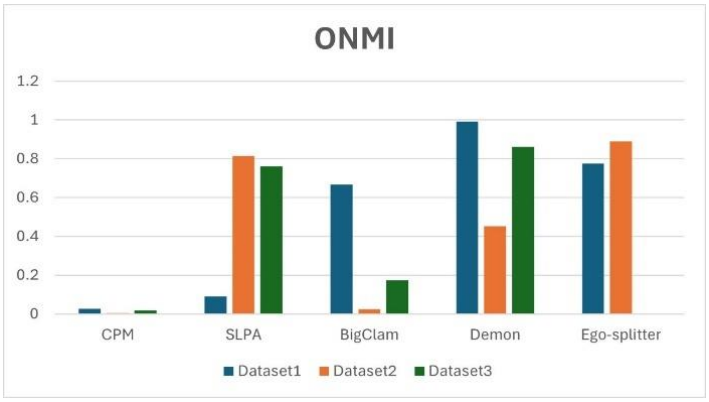

Figure 3.15: Comparison of overlapping normalized mutual information (ONMI)

Figure 3.16: Comparison of extended modularity for overlapping communities

Figure 3.17: Comparison of execution time

3.5 Conclusion

Through the experimental study presented in this chapter, we were able to highlight how the performance of overlapping community detection algorithms varies depending on the properties of the networks used. The results showed that each algorithm has its own relative advantages in certain scenarios, emphasizing the importance of selecting the most suitable algorithm based on the network type and analysis objectives. These findings represent an important step towards a deeper understanding of how to effectively apply these algorithms in real-world social network applications.

General conclusion

Overlapping community detection in social networks is a research area that is still in its developmental and exploratory stages, and it is expected to continue growing in the coming years. This early phase presents many challenges, but it also serves as a strong motivation for developing more accurate and efficient methods and techniques.

The task of detecting overlapping communities is fundamental to understanding the internal structure of social networks and can be formulated as a complex optimization problem. In this research, we conducted a comprehensive comparative study of the most prominent algorithms used for overlapping community detection, evaluating their performance on both real-world and synthetic datasets using standard metrics such as modularity and mutual information.

The experimental results showed that some algorithms exhibit high accuracy and efficiency in detecting overlapping communities, with strong capabilities to handle community overlap effectively. The study also demonstrated that choosing the appropriate algorithm greatly depends on the nature and size of the network, highlighting the importance of accurate evaluation tools and objective comparisons between different approaches.

The analyzed algorithms vary in their implementation complexity some are moderately difficult to apply such as BigClam, while others are characterized by simplicity and scalability such as CPM making them suitable for use in a wide range of practical applications, such as social behavior analysis, recommendation systems, and cybersecurity.

- **Future recommendations**
  - It is recommended to focus on developing overlapping community detection algorithms that can support large and complex networks, especially those with weighted edges that reflect varying strengths of relationships between nodes.
  - It is important to expand the scope of research to include heterogeneous networks that involve multiple types of nodes and edges, thereby enhancing the comprehensiveness and flexibility of the algorithms
  - It is suggested to design recommendation models based on the structure of overlapping communities to improve the accuracy and efficiency of recommendation systems in social networks.

- It is also recommended to incorporate the concept of overlapping communities more deeply into algorithms, enabling support for fuzzy partitions that naturally allow nodes to belong to multiple communities, thus offering a more accurate reflection of social reality.

References

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Structural Analysis in the Social Sciences). Cambridge; New York: Cambridge University Press, 1994, vol. 8, isbn: 0521387078. [Online]. Available: https://books.google.com/books?id=CAm2DpIqRUIC.

[2] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York: Cambridge University Press, 2010, isbn: 9780521195331. [Online]. Available: https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf.

[3] J. van Dijck, *The Culture of Connectivity: A Critical History of Social Media*. New York: Oxford University Press, 2013, isbn: 9780199970780. [Online]. Available: https://books.google.com/books?id=A6BqrWGIaFIC.

[4] The Baylor Lariat, *Social media through the ages: Revolutionizing digital content*, Accessed: Jan. 5, 2025. Available at: https://baylorlariat.com/2018/11/13/social- media- through- the- ages- revolutionizing- digital- content/, 2018.

[5] D. Zahay, M. L. Roberts, J. Parker, D. I. Barker, and M. S. Barker, *Social Media Marketing: A Strategic Approach*, 3rd ed. Cengage Learning, 2023, isbn: 9780357516188. [Online]. Available: https://books.google.com/books?id=mo56CgAAQBAJ.

[6] N. E. Louafi, ``Analyse des réseaux sociaux et détection de communautés," Mémoire de fin d'études, Master en Informatique, Option Systèmes Informatiques, Université 8 Mai 1945 - Guelma, Algérie, 2019.

[7] C. Bahloul, ``Appariement de graphes pondérés : Approche spectrale," Option : Modélisation Informatique des Connaissances et du Raisonnement, Mémoire de Master, Université Dr. Tahar Moulay - Saïda, Algérie, 2019.

[8] S. Kadri. ``Les fondements de la théorie des graphes – chapitre 1: Concepts de base." [Online]. Disponible en ligne., Université Mohamed Boudiaf de M'sila, Faculté des Mathématiques et de l'Informatique, Département d'informatique. (2018), [Online]. Available: https://kadrisaid28.wixsite.com/sgadri.

[9] A. Farouzi, ``Exploration des bases de données orientées graphe : Énumération des triangles dans les graphes à grande échelle," [En ligne]. Disponible : https://theses.hal . science / tel - 04326161, Thèse de doctorat, ESI-Sidi Bel Abbès et ISAE-ENSMA, 2023.

[10] D. B. West, *Introduction to Graph Theory*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2001.

[11] Wolfram MathWorld. ``Empty graph." [Online]. Disponible en ligne. Consulté en mai 2025. (n.d.), [Online]. Available: https://mathworld.wolfram.com/EmptyGraph.html.

[12] E. Chen. ``Math 179: Graph theory, lecture notes." [Online]. Disponible en ligne. Consulté en mai 2025. (n.d.), [Online]. Available: https://web.evanchen.cc/notes/SJSU179.pdf.

[13] Stanford University. ``Relations and graphs, lecture notes." [Online]. Disponible en ligne. Consulté en mai 2025. (n.d.), [Online]. Available: https://web.stanford.edu/class/archive/cs/cs103/cs103.1164/lectures/09/Small09.pdf.

[14] Testbook. ``Antisymmetric relation." [Online]. Disponible en ligne. Consulté en mai 2025. (n.d.), [Online]. Available: https://testbook.com/maths/antisymmetric-relation.

[15] J. D. Dixon and B. Mortimer, ``Arc-transitive graphs," in *Permutation Groups*, Springer, 1996, pp. 130–147.

[16] ProofWiki. ``Multigraph." [Online]. Disponible en ligne. Consulté en mai 2025. (), [Online]. Available: https://proofwiki.org/wiki/Category%3ADefinitions/Multigraphs.

[17] H. Bada, ``Community detection algorithm based on discrete particle swarm optimization," Master's thesis, University of M'Sila, Department of Computer Science, 2024.

[18] N. Lograda, ``La détection de communautés dans les réseaux sociaux," Mémoire de Master, Université Mohamed Boudiaf - M'Sila, Département d'Informatique, 2019.

[19] A.-L. Barabási, *Network Science*. Cambridge, U.K.: Cambridge University Press, 2016, [Online]. Disponible en ligne. Consulté en mai 2025. [Online]. Available: https://networksciencebook.com/chapter/3.

[20] M. E. J. Newman, *Networks: An Introduction*. Oxford, U.K.: Oxford University Press, 2010, [Online]. Disponible en ligne. Consulté en mai 2025. [Online]. Available: https://books.google.dz/books?id=rbxPm93PRY8C.

[21] R. Missaoui and I. Sarr, Eds., *Social Network Analysis: Community Detection and Evolution*. Cham, Switzerland: Springer, 2015, [Online]. Disponible en ligne. Consulté en mai 2025. [Online]. Available: https://books.google.dz/books/edition/Social_Network_Analysis_Community_Detect/_6kqBgAAQBAJ?hl=ar&gbpv=1.

[22] P. Karampelas, J. Kawash, and T. Özyer, Eds., *From Security to Community Detection in Social Networking Platforms*. Cham, Switzerland: Springer, 2019, [Online]. Disponible en ligne. Consulté en mai 2025. [Online]. Available: https : / / www . google.dz/books/edition/From_Security_to_Community_Detection_in/mJiRDwAAQBAJ?hl=ar&gbpv=1.

[23] F. Z. Benhassine, K. Khelif, and H. Maîche, ``A framework for overlapping and non-overlapping communities detection based on seed extension and label propagation," *Physica A: Statistical Mechanics and its Applications*, vol. 660, p. 135 711, 2025. doi: 10.1016/j.physa.2024.135711. [Online]. Available: https://ideas.repec.org/a/eee/phsmap/v660y2025ics0378437125000147.html.

[24] I. Despot, *Understanding community detection algorithms with python networkx*, https://memgraph.com/blog/community-detection-algorithms-with-python-networkx, [Accessed: 21-05-2025], 2021.

[25] Team Statworx, *Community detection with louvain and infomap*, https : / / www . statworx.com/en/content-hub/blog/community-detection-with-louvain-and-infomap, [Accessed: 21-05-2025], 2020.

[26] *Louvain method*, https://en .wikipedia.org/wiki/Louvain_ method, [Accessed: 01-04-2025], 2025.

[27] L. Tang and H. Liu, *Community Detection and Mining in Social Media* (Synthesis Lectures on Data Mining and Knowledge Discovery 1). Morgan & Claypool Publishers, 2010, vol. 2. doi: 10.2200/S00281ED1V01Y200912DMK002. [Online]. Available: https://books.google.dz/books?id=lP2dgtLcdC4C.

[28] Z. Ding, X. Zhang, D. Sun, and B. Luo, ``Overlapping community detection based on network decomposition," *Scientific Reports*, vol. 6, p. 24 115, 2016. doi: 10.1038/srep24115. [Online]. Available: https://www.nature.com/articles/srep24115.

[29] S. Gupta, ``Overlapping community detection in social networks," [Accessed: 21-05-2025], M.S. thesis, San José State University, San José, California, USA, 2011. [Online]. Available: https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=2011&context=etd_projects.

[30] A. Ponomarenko, L. Pitsoulis, and M. Shamshetdinov, ``Overlapping community detection in networks based on link partitioning and partitioning around medoids," *arXiv preprint arXiv:1907.08731*, 2021, [Accessed: 28-05-2025]. [Online]. Available: https://arxiv.org/abs/1907.08731.

[31] F. Ferdowsi and K. A. Samani, ``Detecting overlapping communities in complex networks using non-cooperative games," *Scientific Reports*, vol. 12, p. 11 054, 2022. doi: 10.1038/s41598-022-15095-9. [Online]. Available: https://www.nature.com/articles/s41598-022-15095-9.

[32] S. S. Blankenship and W. E. A. Ruona, *Exploring knowledge sharing among members of a community of practice*, [Accessed: 28-05-2025], 2008. [Online]. Available: https://files.eric.ed.gov/fulltext/ED501645.pdf.

[33] Trainual, *The benefits of sharing knowledge*, [Accessed: 28-05-2025], 2022. [Online]. Available: https://trainual.com/manual/sharing-knowledge.

[34] J. McAuley and J. Leskovec, ``Discovering social circles in ego networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 8, no. 1, 2014. doi: 10.1145/2556612. [Online]. Available: https://cs.stanford.edu/people/jure/pubs/circles-tkdd14.pdf.

[35] ERIS - European Research Institute for Social Work. ``Special interest groups." Accessed: 28-May-2025. (2025), [Online]. Available: https://eris.osu.eu/30569/special-interest-groups/.

[36] N. P. Nguyen, T. N. Dinh, D. T. Nguyen, and M. T. Thai, ``Overlapping community structures and their detection on social networks," in *Proceedings of the 2011 IEEE International Conference on Social Computing (SocialCom)*, 2011, pp. 116–123. doi: 10.1109/SocialCom.2011.23. [Online]. Available: https://www.cise.ufl.edu/~mythai/files/socialcom11.pdf.

[37] J. Yang and J. Leskovec, ``Community-affiliation graph model for overlapping network community detection," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, 2012, pp. 1170–1175. doi: 10.1109/ICDM.2012.38. [Online]. Available: https://cs.stanford.edu/people/jure/pubs/agmfit-icdm12.pdf.

[38] K. Musial, P. Kazienko, and T. Kajdanowicz, ``Content patterns in topic-based overlapping communities," *Entropy*, vol. 16, no. 5, pp. 2938–2972, 2014. doi: 10.3390/e16052938. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4000663/.

[39] Community Tool Box. ``Chapter 5, section 2: Community (locality) development." Accessed: 28-May-2025. (2025), [Online]. Available: https://ctb.ku.edu/en/table-of-contents/assessment/promotion-strategies/community-development/main.

[40] J. Yang, J. McAuley, and J. Leskovec, ``Community detection in networks with node attributes," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020. doi: 10.1038/s41467-020-17779-0. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7338183/.

[41] R. Gao, S. Li, X. Shi, Y. Liang, and D. Xu, ``Overlapping community detection based on membership degree propagation," *Entropy*, vol. 23, no. 1, p. 15, 2021. doi: 10.3390/e23010015. [Online]. Available: https://www.mdpi.com/1099-4300/23/1/15.

[42] Z. Wang, W. Zhang, and X. Zhu, ``Multi-objective optimization for overlapping community detection," in *Proceedings of the 2013 International Conference on Social Computing*, 2013, pp. 1–8. [Online]. Available: http://shichuan.org/doc/13.pdf.

[43] J. Yang and J. Leskovec, ``Overlapping community detection at scale: A nonnegative matrix factorization approach," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM)*, 2013, pp. 587–596. doi: 10.1145/2433396.2433471. [Online]. Available: http://i.stanford.edu/~crucis/pubs/paper-nmfagm.pdf.

[44] O. Shchur and S. Günnemann, ``Overlapping community detection with graph neural networks," in *Proceedings of the 1st International Workshop on Deep Learning for Graphs (DLG'19)*, 2019, pp. 1–7. [Online]. Available: https://arxiv.org/pdf/1909.12201.

[45] Y. Wang and X. Liu, ``An influence-based label propagation algorithm for overlapping community detection," *Mathematics*, vol. 11, no. 9, p. 2133, 2023. doi: 10.3390/math11092133. [Online]. Available: https://www.mdpi.com/2227-7390/11/9/2133.

[46] H. A. Jalab, R. S. Al-Atroshi, and A. A. Al-Atroshi, ``Overlapping community detection using multi-objective approach based on rough clustering," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–13, 2020. doi: 10.1155/2020/8851425. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7338163/.

[47] V. Nicosia, G. Mangioni, V. Carchiolo, and M. Malgeri, ``Extending the definition of modularity to directed graphs with overlapping communities," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 03, P03024, 2009. doi: 10.1088/1742-5468/2009/03/P03024. [Online]. Available: https://arxiv.org/abs/0801.1647.

[48] A. F. McDaid, D. Greene, and N. Hurley, ``Normalized mutual information to evaluate overlapping community finding algorithms," *arXiv preprint arXiv:1110.2515*, 2011. [Online]. Available: https://arxiv.org/abs/1110.2515.

[49] J. Xie, B. K. Szymanski, and X. Liu, ``Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," *arXiv preprint arXiv:1109.5720*, 2011. [Online]. Available: https://arxiv.org/abs/1109.5720.

[50] A. Baudin, M. Danisch, S. Kirgizov, C. Magnien, and M. Ghanem, ``Clique percolation method: Memory efficient almost exact communities," in *Proceedings of the 2023 International Conference on Network Science*, 2023. [Online]. Available: https://kirgizov.link/publications/BDKMG.pdf.

[51] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, ``Demon: A local-first discovery method for overlapping communities," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2012, pp. 615–623. doi: 10.1145/2339530.2339630. [Online]. Available: https://arxiv.org/abs/1206.0629.

[52] A. Epasto, S. Lattanzi, and R. Paes Leme, ``Ego-splitting framework: From non-overlapping to overlapping clusters," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 145–154. doi: 10.1145/3097983.3098054. [Online]. Available: https://dl.acm.org/doi/10.1145/3097983.3098054.

[53] Y. Ke, *Slpa-py: Implementation of slpa in python*, https://github.com/YipingNUS/slpa-py/blob/master/karate.txt, Accessed: 03-06-2025, 2013.

[54] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, ``The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396–405, 2003. doi: 10.1007/s00265-003-0651-y. [Online]. Available: https://doi.org/10.1007/s00265-003-0651-y.

[55] V. Krebs, *Books about us politics network*, https://github.com/graphstream/gs-gephi/blob/master/data/polbooks.gml, Accessed: 03-06-2025, 2004.

[56] J. McAuley and J. Leskovec, ``Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 539–547. [Online]. Available: https://snap.stanford.edu/data/egonets-Facebook.html.

[57] V. Krebs, *Les misérables character co-occurrence network*, http://konect.cc/networks/moreno_lesmis/, Accessed: 03-06-2025, 2004.

[58] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Benchmark graphs for testing community detection algorithms*, https://paperswithcode.com/dataset/a-collection-of-lfr-benchmark-graphs, Accessed: 03-06-2025, 2008.

ملخص

يمكن تمثيل العديد من الأنظمة المعقدة في العالم الحقيقي على شكل شبكات اجتماعية معقدة، والتي تكشف عن خصائص هامة مثل وجود المجتمعات المتداخلة. تعد المجتمعات المتداخلة من الظواهر الأساسية التي تعكس التداخل والانتماءات المتعددة للعناصر داخل الشبكة، ولها تطبيقات واسعة في مجالات متعددة مثل تحليل السلوك الاجتماعي، التوصية، والامن السيبراني. في هذا البحث، نقدم دراسة مقارنة شاملة لخوارزميات الكشف عن المجتمعات المتداخلة، مع توضيح مفاهيمها، تصنيفاتها، وتحدياتها. كما نجري تقييما تجريبيا لأداء ابرز هذه الخوارزميات على مجموعات بيانات حقيقية و مولدة، باستخدام مقاييس تقييم معيارية. تهدف الدراسة الى تقديم توصيات تساعد في اختيار الخوارزمية الأنسب حسب طبيعة الشبكة والتطبيق المطلوب.

الكلمات المفتاحية: المجتمعات المتداخلة، الكشف عن المجتمعات، الشبكات الاجتماعية، نظرية الرسوم البيانية، الخوارزميات، التقييم المقارن

Abstract

Many real-world complex systems can be modeled as social networks, which reveal important features such as overlapping communities. Overlapping communities reflect the multiple memberships and interconnections of elements within the network, with wide applications in social behavior analysis, recommendation systems, and cybersecurity. This research provides a comprehensive comparative study of overlapping community detection algorithms, clarifying their concepts, classifications, and challenges. We conduct experimental evaluations of prominent algorithms on real and synthetic datasets using standard evaluation metrics. The study aims to offer recommendations for choosing the best algorithm depending on the application context and network type.

Keywords: Overlapping Communities, Community Detection, Social Networks, Graph Theory, Algorithms, Comparative Evaluation.

Résumé

De nombreux systéms complexes du monde réel peuvent etre représentés sous forme de réseaux sociaux montrant des communautés. Chevauchantes, reflétant les appurtenances multiples entre les éléments. Cette étude compare les algorithmes de detection de communautés chevauchantes, en expliqaunt leur concepts, classifications et défis, avec des experiences sur des données réelles et synthétiques a l'aide de métriques d'évaluation standards. L'objectif est de fournir des recommandations pour choisir l'algorithme le plus adapté en fonction du type de réseau et du contexte d'application.

Mots-clés: Communautés chevauchantes, Detection de Communautés, Réseaux sociaux, Théorie des graphes, Algorithmes, Evaluation Comparative.