“Social Media Community Detection using Machine Learning-based Clustering Analysis “

A

Thesis

Submitted towards the Requirement for the Award of Degree of

Doctor of Philosophy

In

COMPUTER SCIENCE ENGINEERING

Under the Faculty of Engineering and Technology

By

RAVINDRA SINGH YADAV

(Enrollment No- 161588517256)

Under the Supervision of

Dr. BALVEER SINGH

Professor

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING &IT

Year - 2024

P.K. University Shivpuri M.P. -473665

"Social Media Community Detection using Machine Learning-based Clustering Analysis "

A

Thesis

Submitted towards the Requirement for the Award of Degree of

Doctor of Philosophy

In

COMPUTER SCIENCE ENGINEERING

Under the Faculty of Engineering and Technology

By

RAVINDRA SINGH YADAV

(Enrollment No- 161588517256)

Under the Supervision of

Dr. BALVEER SINGH

Professor

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING &IT

Year - 2024

P.K. University Shivpuri M.P. -473665

Signature of Supervisor

P.K.University
Shivpuri (M.P.)

UNIVERSITY ESTABLISED UNDER SECTION 25 OF UGC ACT 1956 VIDE MP GOVERNMENT ACT NO 17 OF 2015 ADRESS VILLEGE THANARA, TEHSIL KARERA,NH27,DISTRICT:SHIVPURI,MP

CERTIFICATE OF THE SUPERVISOR

This is to certify that the work entitled **"Social Media Community Detection sing Machine Learning-based Clustering Analysis"** is a piece of research Work done by Mr. Ravindra Singh Yadav Under My/Our Guidance and supervision for the degree of Doctor of Philosophy of Computer Science Engineering to be Awarded by P. k University (M.P) India.

I certify that the candidate has put an attendance of more than 240 day

with me. To the best of my knowledge and belief the thesis.

I – Embodies the work of the candidate himself/herself.

II – Has duly been completed.

III – Fulfill the requirement of the ordinance relating to the Ph.D. degree of the University.

Signature of the Supervisor

P.K.University
Shivpuri (M.P.)

UNIVERSITY ESTABLISED UNDER SECTION 25 OF UGC ACT 1956 VIDE MP GOVERNMENT ACT NO 17 OF 2015 ADRESS VILLEGE THANARA, TEHSIL KARERA,NH27,DISTRICT:SHIVPURI,MP

DECLARATION BY THE CANDIDATE

I declare that the thesis entitled **"Social Media Community Detection sing Machine Learning-based Clustering Analysis"** Is my own work conducted under the supervision of Dr. Balveer Singh Approved by Research Degree Committee of the university and that I have put more than 240 days of attendance withSupervisor at the center.

I further declare that to the best of my knowledge the thesis does not contain my part of any work with Has been submitted for the award of any degree either by this University or by any other University Without proper citation.

Signature of the candidate

Date: ……………………………
Place: ……………………………..

P.K.University
Shivpuri (M.P.)

UNIVERSITY ESTABLISED UNDER SECTION 25 OF UGC ACT 1956 VIDE MP GOVERNMENT ACT NO 17 OF 2015 ADRESS VILLEGE THANARA, TEHSIL KARERA, NH27, DISTRICT: SHIVPURI, MP

FORWARDING LETTER OF HEAD OF INSTITUTION

The Ph.D. thesis entitled **"Social Media Community Detection sing Machine Learning-based Clustering Analysis"** Submitted by Mr. Ravindra Singh Yadav Is forwarded to the university in six copies. The candidate has paid the necessary fees and there are no duesoutstanding against him/her.

Name Ravindra Singh Yadav

Seal

Date:

Place:

(Signature of Head of Institution where theCandidate was registered for Ph.D. degree)

Signature of the Supervisor Date:

Date:

Place:

Address:

ACKNOWLEDGEMENT

First and foremost, Praises and thanks to the God for his showers of abundant blessings throughout my research work to complete the research successfully

It is my pleasure to express my thanks to P.K. University Hon'ble Chancellor Mr. J. P. Sharma, Hon'ble Vice Chancellor Prof. G.Pavan Kumar for their cooperation and also for giving me the opportunity to do the research work at their prestigious University.

I would like to express my sincere thanks to Dr. Aiman Fatma, Dean Academics. My special thanks go to Dr. Nalla Bhaskar, Dean Research and I/c Ph.D. Cell and his Office Assistant Mr. Ashish Gupta, Ms Megha Sharma.

I would like to express my deep and sincere gratitude to my Research Supervisor. Dr. Balveer Singh, for giving me the opportunity to do research and providing invaluable guidance throughout this research. His immense knowledge, intellectual thinking, vision. dynamism and motivation have deeply inspired me. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under her guidance.

I like to thank Dr Deepesh Namdev, Registrar, Ms. Nisha Yadav, Librarian and Mr. Kamlesh Yadav, IT. Cell.

I sincerely thank my loving parents who has always been there with me in every situation. Last but not the least, I am extremely thankful to my beloved Mrs. Namrata Yadav & my children Manya and Khayati for their moral support, patience and encouragement throughout the research work. Above all, I thanks to Almighty for granting me health, strength & the wisdom to commence this research work and enabling me to its completion

Ravindra Singh

Research Scholar
Ravindra Singh Yadav
(Enrollment No. 161588517256)

P. K. UNIVERSITY
SHIVPURI (M.P.)

University Established Under section 2(F) of UGC ACT 1956 Vide MP Government Act No 17 of 2015

ef. No. PKU/06/10/2018/Engg/CSE/01 Date 04/08/2018

Course Work Certificate

To

Ravindra Singh Yadav.

Dear Student,

This is to certify that **Mr.Ravindra Singh Yadav** Enrollment no:**161588517256** S/o Mr. Jaswant Singh Yadav, student of Ph.D. **(Computer Science and Engineering)** under the **Faculty of Engineering** and Technology has successfully passed the course work examination with **'A'** from P.K.University, Karera Shivpuri.

Registrar

P.K.UNIVERSITY
SHIVPURI (M.P.)

UniversityEstablished Under section 2F of UGC ACT 1956 Vide MP Government Act No 17 of 2015

CENTRAL LIBRARY

Ref. No. PKU/ C.LIB /2024/PLAG. CERT./132 **Date: 27.02.2024**

CERTIFICATE OF PLAGIARISM REPORT

1. Name of the Research Scholar	:	Ravindra Singh Yadav
2. Course of Study	:	Doctor of Philosophy (Ph.D.)
3. Title of the Thesis	:	Social Media Community Detection using Machine Learning-based Clustering Analysis
4. Name of the Supervisor	:	Dr. Balveer Singh
5. Department	:	Computer Science & Engineering
6. Subject	:	Computer Science & Engineering
7. Acceptable Maximum Limit	:	10% (As per UGC Norms)
8. Percentage of Similarity of Contents Identified	:	1%
9. Software Used	:	Ouriginal (Formerly URKUND)
10. Date of Verification	:	19.01.2024

Nadeem 27.02.24

Signature of Ouriginal Coordinator
(Librarian, Central Library)
P.K.University, Shivpuri (M.P.)

ADD: VIL: THANRA, TEHSIL: KARERA, NH-27, DIST: SHIVPURI (M.P.) -473665
MOB: 7241115902, Email: library.pku@gmail.com

P.K.University
Shivpuri (M.P.)

UNIVERSITY ESTABLISED UNDER SECTION 25 OF UGC ACT 1956 VIDE MP GOVERNMENT ACT NO 17 OF 2015 ADRESS VILLEGE THANARA, TEHSIL KARERA, NH27, DISTRICT: SHIVPURI, MP

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: entitled **"Social Media Community Detection sing Machine Learning-based Clustering Analysis"** Submitted by Mr. Ravindra Singh Yadav

COPYRIGHT TRANSFER

The undersigned hereby assigns to the P.K. University, Shivpuri all copyrights that exist in and for the above thesis submitted for the award of the Ph.D. degree.

Date:

Ravindra Singh

Name of the scholar

Ravindra Singh Yadav

Note: However, the author may reproduce/publish or authorize others to reproduce, material extracted verbatim from the thesis or derivative of the thesis for authors personal use provided that the source and the University's copyright notice are indicated.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A platform for collective viewpoint for online marketing, advertising, political campaigns, and other purposes is provided by social media podium. It organizes a community of like-minded end users above the explicit group. The communal group on social media is the structure of the community. of widely dispersed individuals with shared interests in a product, community issue, or any other axis.

The intrinsic community aspects of social media stem from the implicit and explicit characteristics of their users. The graphical form of social media makes it easy to identify the explicit nature of both active and passive users. Conversely, end users' engagement determines how much implicit feature information is included. While integrating the implicit properties of quiet and inactive users across the community is a laborious task, the implicit traits of often active users are widely available.

The goal of this effort is to substantially bind passive users throughout the community by developing an analytical and methodological framework for community detection. Additionally, this paper presents the idea of an unsupervised machine learning technique from a social media network perspective. In order to maximize influence, this idea aids in determining the trade-off between the density of connections and the similarity of node properties. Three distinct modules were sculpted into the created framework for community detection.

Semantic relation-based modularity-optimized community discovery technique for heterogeneous networks is covered in the first module. By combining the network's content analysis with link analysis, this effort seeks to improve the network's modularity value. As a result, the network was constructed using the similarity values between individuals' shares as indirect linkages. The approaches for content and link-based clustering that are being discussed employ a greedy hierarchical clustering algorithm that leverages indirect connections with the network structure to prioritize the topological and semantic grouping of the nodes that are closest to each other. This study compares the effects of semantic relationships on two optimization algorithms—the Parliamentary Optimization Algorithm (POA) and the Modularity Optimization Algorithm (MOA)—for the purpose of community discovery. In the end, the work's modularity

and NMI were assessed using six actual, heterogeneous network data sets, and the resulting informative community showed a sufficient modularity rate.

Vi A comparative examination of the overlapping community detection approach is presented in the second module. This lesson provides a thorough analysis of single and multi-purpose functions for community discovery, as well as a comparative study of heuristic overlap community detection algorithms via social media. In the social media data set, this module saw that the performance rate of the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO varied with network density. It achieved a greater performance rate in the dense ACF network and a significantly lower one in the weakly dense WA data set. Additionally, over a denser network, TLBO and ICA are able to extract a greater informative community. Simultaneously, the less dense networks provide superior outcomes for SEOA and HSA.

In contrast, the third article in this series offered community detection methods based on Parliamentary Optimization (MPPOA) and Multi Purpose functions. For the first set of data, the Parliamentary Optimization Algorithm (POA) population was constructed in a Python environment. Following the division of the population into a predetermined number of groups, the power values of each group were computed. Vulnerable groups are removed from the population based on the specified deletion probability, whilst strong groups exhibit joining in accordance with the given combination probability value. The problem's outcome has been neared. The program steps were repeated until all groups were combined and the algorithm's termination condition was satisfied. The final stage's individual with the highest eligibility values among the remaining data was approved as the algorithm's solution to the overlapping community discovery problem. Performs a study that has been presented and assessed using four actual and one artificial network-based social media data sets after that. To determine the impact of Single and Multi-Purpose Function on community detection performance, a feature assessment comparison is conducted. This study concludes with a comparison of the MPPOA algorithm with three heuristic overlap community discovery algorithms across six authentic social media data sets.

Keywords : overlap community identification, influence maximization, social theories, social media, overlap community detection, and legislative optimization

Chapter 1
Introduction

Social media is becoming a more popular topic for scholars in the modern world. Social media users create enormous amounts of data. There on several mining tasks involved in social media mining order t preserve the user-generated data. There are plenty of them. Social media platforms where user create their own communities based on their interests. It's well known that social is a vast virtual environment since so many people use it to create profiles and connect with all kinds of organizations. Understanding the user's past is essential to their behavior. cause it difficult to discern a single user's activity on a social network, community detection is required.

Popular websites social media exist where users may form various social connections and groups individuals with varying opinions and points of view. social networking sites. Here, people can exchange materials and resources. In 2011, Jack and Scott invented the term "social media" to refer to online social networks. Some examples of these sites include (https://www.facebook.com), (https://www.twitter.com), (https://www.friendsnet.com), and so on. The grouping collectively known as social media gives users the opportunity. Oxford University defined media in 2020 as a web-based tool networking [1].

Social networking is th process of interacting with other users on specific social media platforms or locating individuals who share interests. According to the organization for economic operation and development, Smith's 2020 research revealed that in order for content provided by users to be considered UGC (user-generated content), it must fulfill three fundamental requirements. 1.) All web users or a specific group must see the content (mails and instant messages excluded). 2.) It must be imaginative and unique rather than a copy of another user's work. It should not be used for commercial purposes and should be made independently of professional routines.

According to Smith et al., the daily creation of new websites makes it challenging to categorize social media a systematic way. This thesis presents social media ass new topic of study for researchers [1]. The practice of mining social media a procedure for analyzing, identifying, and extracting relevant patterns from social networks [3]. The have combined with computer techniques through social media dig.

When examining vast amounts of social media data, social media mining clarifies fundamental ideas and concepts., and so on have all been discussed in this mining.

They offer a comprehensive set of tools for social media mining that enable the formal representation, modeling, measurement, and extraction of significant patterns from vast social media networks. Social media platforms produce user data that differs from conventional attribute-values of information for mining Greek data. The data created by social media platforms is often loud, dispersed, ill-structured, and frequent. Because social media data has so many unique qualities, data mining tasks are difficult, necessitating the development novel approaches and algorithms[3].

1.1 Social Media

media is becoming the must effective technique to communicate with users of the internet and exchange information. The term "social media" refers to a collection of online tools and platforms that let people engage in social networks and produce and distribute content.[1]. Social media is a "platform for creating information, forging clear and constructive relationships," according to Bruglieri et al. [2]. Web technologies are made available by to share user information and content with a hugs audience.

Numerous globally renowned platforms enable users to participate in diverse social exchanges with groups that have varying perspectives and viewpoints. Online dating sites are what these websites are called. Users able too exchange material and resources. Social networking sites include, but no limited , Facebook, Twitter, and Net Friends. Using social networking sites or finding share your interests are two extra. social networking. According to Moscato et al. [3], user material on need to accomplish the following goals:

1. The content should be sent right away via email and message to every user on the website the assigned team.
2. Avoid using other people's content and instead be creative.
3. It must be developed without any technological problems and not be put to use for profit.

Social media was described by Hafiene et al. [4] using the following seven construction pieces as their framework:

(a.) Identity: The user must their personal Infor. in this area, including: name, gender, age, occupation, place of residence, and any information the user choose to provide.

(b.) Conversions: Through social media, end users may no connect with one another based on common interests and occupations. Numerous websites offer solutions for group and individual communication. Distractions like tweets, blogs, the ability to similar interests, the possibility for users to find love, etc., are all good reasons to provide them.
(c) Sharing: Social language encompasses the idea of sharing. By exchanging ideas, 1 may bring about transformation. Users are able to send and receive material with this tablet. Sharing throughout time necessitates adjustment between parties. Because of this, users are able to communicate and connect with one another.
(d.) Presence: This option displays the size of the user to determine whether another user is accessible. Knowing other users' whereabouts in the visible world is part of this.
(e.)Relationship: On social media, communication is a representation of user engagement. Social media interactions can in a variety of forms, such as fan favorite, casual, or a user group that shares and discusses a subject.
(f) Reputation: Due to social media, people might have +ve or -negative reputation depending on whether they network site status. The majority of the work is completed in this area, but because it is an accessible environment, it is exceedingly challenging to earn people' complete confidence.
(g) Groups Identity: The presence of groups indicates a low level of user interaction.
A significant portion was the group and society rely on friends, connections, followers, and the most essential items.

1.2 Social Media Mining

The goal of using social media i to model, assess, and illustrate the many applications of [5]. The terms "social media mining" refer to the ideas and methods using social media data study. A rang of instruments are used inn social media mining to lawfully For big media networks, create, rate, and utilize existing media, for instance. User data produce by media platforms differs from that of conventional mining resources. Social media generates a lot of stuff that is spread widely and without context. Numerous issues arise throughout the mining process as a result of all these data medium properties.
The following issues are frequently present while mining social media [6]:

1.Social media data volume: The amount of material available on social media platforms is enormous. There must be a legitimate need for each individual if we are to respect their rights. There is little data available per individual 4specific Individual. Thus, we need to capitalize on the influencer.

2. Data extraction: Programming applications may gather quantifiable data for everyday needs by utilizing the API, which handles social may be accessed using certain methods, however data access is not unlimited.

3. Noise reduction: The data media offers an outstanding level of noise reduction.

As such, one of the hardest things for mining to do is to eliminate noise from data. The vital information could be neglected in this situation, thus the term cannot be muted. Here, what constitutes sound relies on the task that we perform.

Table 1.1: mining task

User based	Relation based	Content based
Spammer Detection	Strength prediction	Sentiment1 Analysis
User classification	Social1 Tie Prediction	Feature Selection
Community Detection	Link Prediction	Recommendation

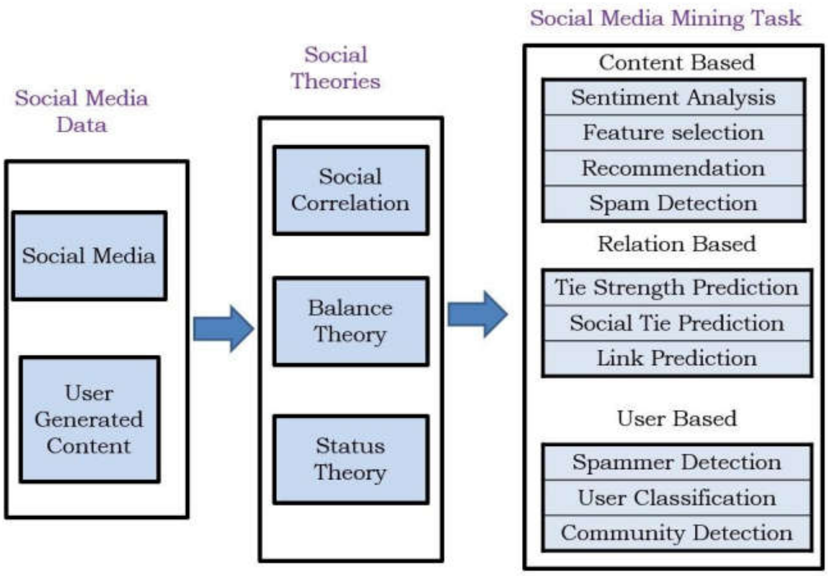


Figure 1.1: Description Hierarchical of Social theories

1.2 Social theories

Situational memory, social change, and balanced cognition are the three categories of social theories. Figure 1.1 depicts a hierarchical explanation of social theory [7].

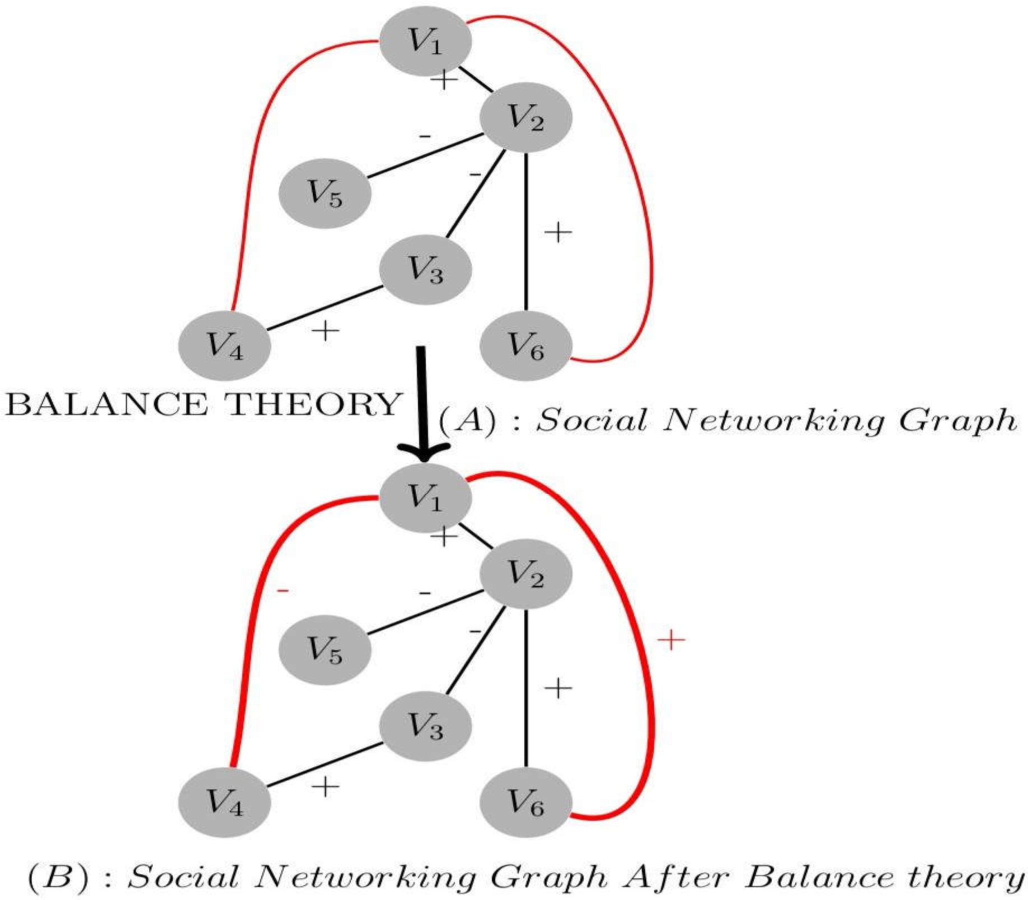


Figure 1.2: Social Balance Theory

1.3.1 Balance Theory

The idea of social inequality serves as an example of the imbalance that occurs in interpersonal connections. Assume that there are 2 components, v1 and v2, that correspond to the right symbols. This indicates that if the sharp mark is off, v1 nv2 are buddies. They are not connected.

1.3.2 Status Theory's

The typical interactions between neighbors and people are reflected in public concepts. This may be summed up in a single illustration. According to him, if v1 is low andv2 is less than v3, then v1 ought to be lower thanv6. The places are represented by the numerals, one.

symbolizes the status of virtuous individuals. and the incorrect ones. The functional output of code Y X shows that, in the same way that corners and dotted lines reflect all relationships, Y is greater than X.

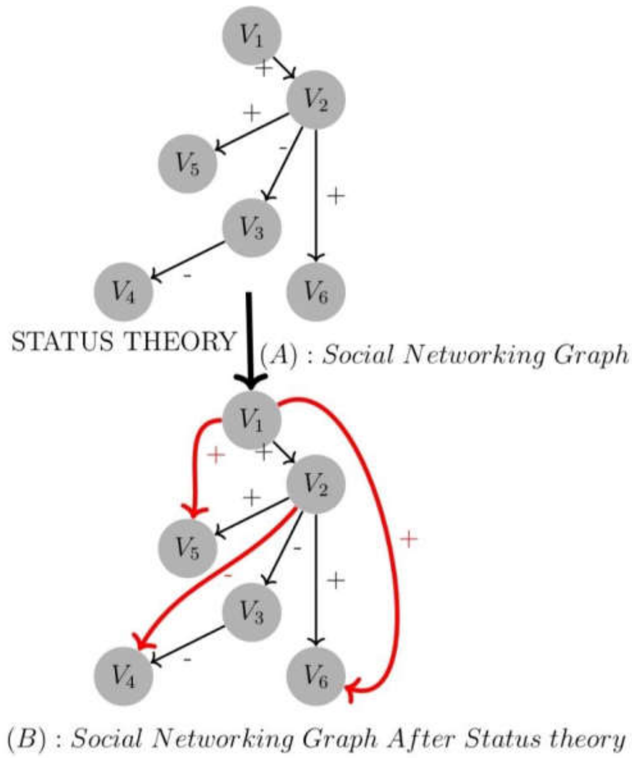


Figure 1.3: Social Status Theory

1.3.3 Social Correlations

Social network behavior and social networks are connected. Hemophilia, mixing, and labor are the three main areas into which public relations may be categorized.

1.3.4 Influence

Inspiration is a work of art or behavior that has an online impact on someone else. In a particular network, Figure 1.4(c) depicts a node's order to another user. Measures that quantify the effect between the persons may be classified into two categories: prediction-based and observation-driven.

(a) Prediction-based: It is assumed that the individual in question have strength or not. The network's location and structure re taken into consideration while making the decision. For instance, it's believed that the quantity of pals with whom one may discuss the influence users have on society. Degree information is utilized many times on Twitter networks to identify possible users.

(b) Observation-based: Using this kind of evaluation, we assess users' power according to their level of perplexity. A person can change depending in the circumstances. Three distinct scenarios that have varying effects on one another are explained below.

-When individuals are held up as models: the fashion industry, celebrities, and educators are the ones that cultivate this type of personality. The degree of lead owing to celebrity or fashion determines the size of the audience. These figures can be used as a clear impact.

- When people discuss information: The scenario in this instance is predicated on the quantity of anti-virus offenses. It also relies on the user's acceptance rate.

- The product's value rises with user participation: the cost of the product rises in proportion to this kind of feedback. Additionally, the product's price to sales ratio evaluates the product's influence.

1.3.4 Homophily

A method of linking users based on great resemblance is called homophily. The same users are associated to the same objects in Figure 1.4(B), which is a classic example of homology.

1.4 Community Detection

Locating communities inside a social network is known as community detection. Groups, clusters, and coherent subgroups are other names for communities. Humans by nature create communities and groupings according to their shared interests and traits. In order to comprehend better understand the intricate social network, we must determine how many communities allow us to split the network into smaller groups or clusters, as seen in figure

A number of queries about community detection come up. Among the questions are the following: the first concerns community detection, the second concerns community evolution, and the third concerns assessment metrics for communities that have been identified. Anytime We identify communities based on certain members or types of communities [3].

1.5 Community Detection Process

Both member-based and group based community discovery algorithms fall under this category. group-based community detection s based on the group's characteristics, whereas member-based community discovery is based other.

picture 1 illustrates the user's interest [3].6. Member-based and Group-based are the two categories into which the community discovery technique falls. There are more subcategories with on these two main groups.

1.6 Member Basic Community Detection

Since it is believed that comparable members would belong communities, member-based community detection is carried out based the member's attributes. The nodes that form a cycle in a graph or network are regarded as community because They have a tight relationship. Node degree, n similarity, and node reachability are the three properties that determine whether a subgraph of a graph is a community.

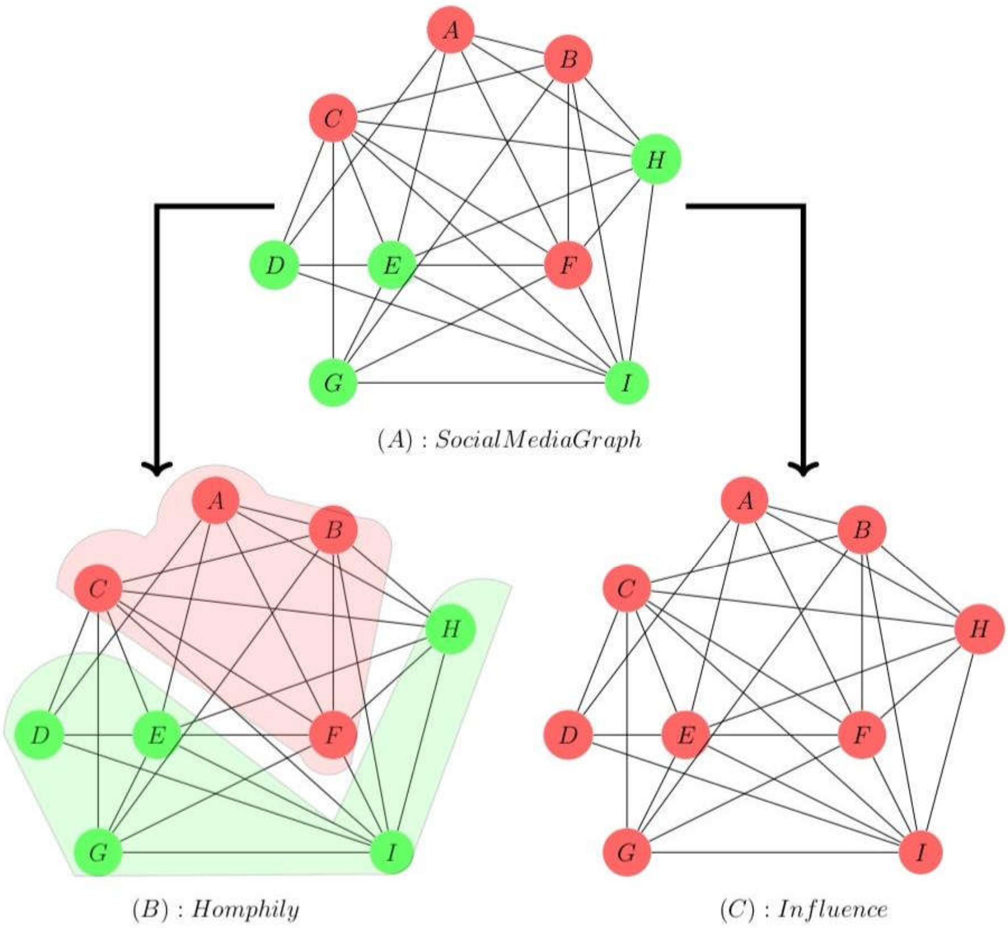


Figure 1.4: Social Correlation Theory

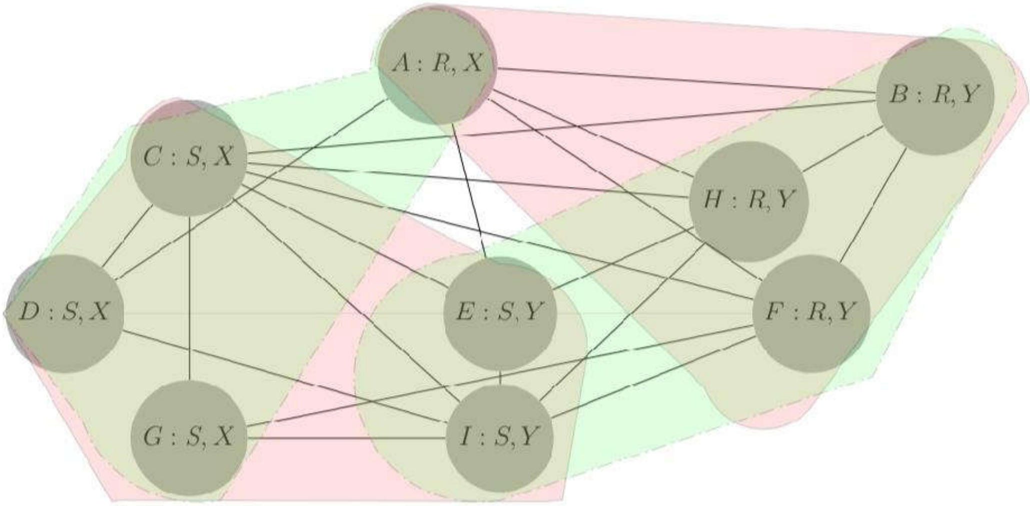


Figure 1.5: Community over Social Media

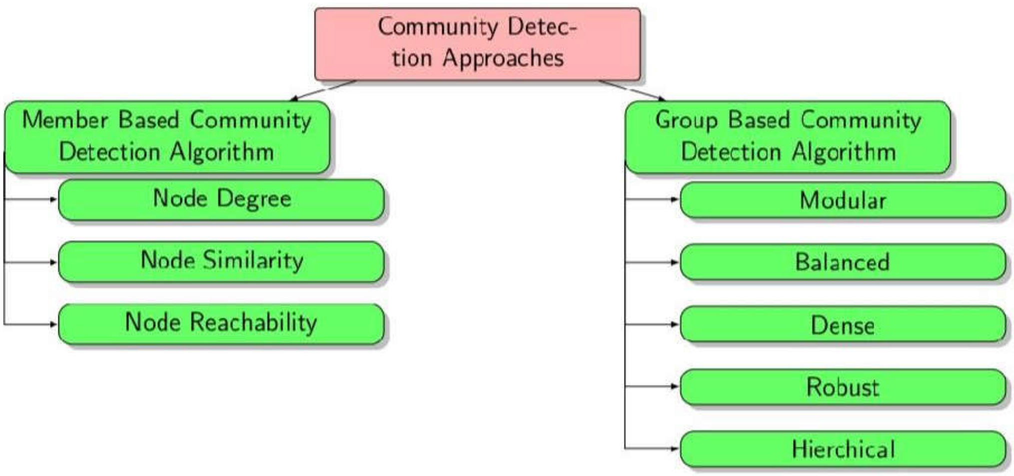


Figure 1.6: Community Detection Algorithm Hierarchy

(a) Node Degree: Node degree, commonly referred to as clique, is the basis for subgraph searching in a network. A clique is a fully linked subgraph in which every node and every pair of nodes is connected. Assume the size of the clique is

A subgraph has k nodes, each of which induces a node degree of K-1. The Brute-force clique identification technique and the clique-percolation method are the two algorithms for identifying cliques inside a network or graph.

(b) Node Reachability: Reachability denotes the ability of two node pairs to communicate with one another. When we discuss node reachable, we search for subgraph in nodes may be reached by paths to other nodes. Two extreme locations for When nodes are intended to bin the Same community, reachability is attained.

There are 2requirements for node reachability: 1) they must be adjacent enough considered immediate neighbors; and 2) there must be a path connecting them, regardless of distance. We can find related components using any traversal strategy justify the first property.

(c) Node Similarity: This refers to the method of determining how same two nodes to one another.

Two are considered to be part the same community if their similarities are sufficient.

Following the determination the nodes' similarities, classical algorithm, the settlements are located. Similarity in structural equivalency is measured or determined based on neighbor.

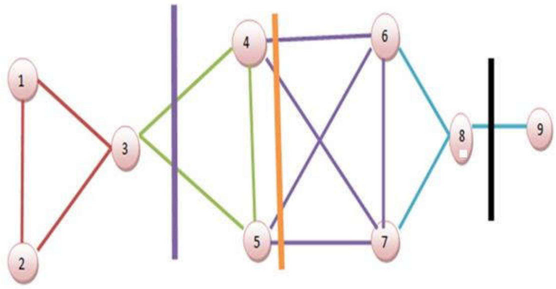
$O(vi, vj) = /N(vi) \quad N(vj) \quad /$ (1.1)

Because are more neighbors in a vast network, the value the equations grows quickly. There are two distinct normalizing methods: cosine similarity and Jaccard similarity. Communities that overlap can alter the node's similarity score.

1.7 Group Basic Community Detection

In group-based community detection, the group's characteristics are taken into account. The following communities fall under this category of community detection: balanced, modular, resilient, dense, and hierarchical communities.

Figure 1.7: Minimum-cut method



(a.) Balanced Communities: They divided the network into many divisions, each of which is represented as a community. They did this by using graph-based clustering [3], which been shown to be the effective method for finding the communities. Here they

have employed the minimum-cut technique in figure 1.7, to identify balancing communities. It's among the earliest techniques for partitioning a network into many sections. Its primary purpose is load balancing, which reduces the amount of communication between CPU nodes. This approach divides the network into a predetermined number of about equal-sized sections. Occasionally, communities with a single node are produced when the minimum-cut approach is used to find communities on a network.

(a) Robust Communities: In robust communities, the goal is to identify a subgraph that remains connected even if a single edge or node is eliminated. Assume that m is the minimal number in an m-vertex graph.

of nodes that need to be eliminated in to break the graph. indicates that there are least m separate pathways connecting any two nodes. m-edge the graph is a comparable subgraph which disconnecting the graph requires removing at lem edges.

(c) Modular Communities: Communities that identify based on modularity are referred to as modular communities. The community have random structure. They've examined a undirected grap G(v, e), with degree = |e| = m.

edges are unknown, but the node's is known. Examine two nodes, Va and Vb, that have relative degrees of da and db. The anticipated number edge between these nodes must be ascertained. At this point, there di 2m chance that an edge will leave node Va and join to Vb.

(d) Dense Communities: Dense are ones where their p lot of interaction. Communities of this kind are developed based on shared interests. Some properties are there to help determine the community's density.

Within the dense community are cliques, clans, and clubs. A dense community is a fully linked or interconnected subgraph or clique.

(e) Hierarchical Communities: In a hierarchical community, each subcommunity and super community is independent of the other. The hierarchical clustering technique, which considers n nodes to be either a community or one, has been employed by them. In this

The graph or network is initially divided according n number of nodes using the clustering technique.

Subsequently, the neighboring communities are merged into a single new community, and the to number f communities is eventually revealed.

1.7 Motivation

In the current era of digital marketing, advertising, and PR, social media platforms provide a dynamic viewpoint for categorizing people and customers who have similar interests through community detection. Social media community is the worldwide cooperative association. disperse people with like interests in a shared subject or item. An algorithm for detecting communities on social media groups together who are very active around a common problem.

One of the newest areas of social media mining is community detection. Scalability and community quality are major problems with community detection Researchers Windy have recently discovered communities in social media by using clustering and media mining method. However, the performance of research does not yield meaningful findings because of a lack of network information.

1.8 Objective

The goal is to extract relevant characteristics far social media date that illustrate intricate relationships and help discover key communities. Specifically, this includes:

(a) Providing an overview of unsupervised machine learn technique using graph theory

influence maximization from a social media viewpoint to determine the trade-off between node attribute similarity and connection density.

(b) Social atom clustering for efficient community identification.

(c) Use the impact node the starting point f community discovery to greatly increase the clustering performance.

1.10 Research Envisaged Planning of Work

By merging graphical and social theory-based features, as well identifying the social media community's structure, this study proposes an impact maximization framework to considerably bind end-users intrinsic traits. The structure that is being shown comprises four distinct modules for the purpose of community detection.

A structured computational framework for assessing the performance of the benchmark method over six actual network data sets is presented in the preliminary work of this paper. generates information that is overlapping on the strengths a weaknesses of community organizing through social media.

graphical based community detection framework (GCDF) is presented in the second module. When GCDF first finds a clique structure as seed community, it clusters the homogenous data to establish the desi community structure.

node above the node v feature focused on the group. conducts a comparison study of group-centric and node-centric features across six basic network-based social media data sets after that. To determine the impact of the graphical feature set on assessment is conducted.

An analytical farm for assessing the impact or social theory oj community detection is presented in the third module. First, this approach uses correlation theory, social balance, and status to extract implicit node information that improves.

In contrast, the fourth module creates a framework for centralized influence maximization (IMCD) that is use organize communities on social media. Init, IMCDs combines node-centric data to find constrained clique structure (seed community).

and feature that is focused on the group. Social theory then gives a more constrained cleeque structure, shows intra-cleeque node , and broadens the definition oy clique structure. Lastly, the consequent community structure og the spindle or the highest influence node is found via IMCD.

The results papers might b used to investigate end users' social media community activities. It was then applied to identify terrorist organizations on social media as well research, educational, and e-commerce groups.

1.11 Organization of this Thesis

This thesis divided into the six chapters that follow.

(i) The "Introduction" chapter of Chapter 1 explains the basics and importance of social uu community detection. information presented in chapter serves as the main justification for the idea that appropriate organization .

social media community can improve the experience of keeping an eye activities. Social and social media mining are covered first, followed by an example of community detection.

It incorporates a component of at the same are time, such as correlation theory, social balance, and status. Additionally, a summary of the thesis' goal and contribution to the understanding of disjoint community identification on social media is provided.

(ii) "Background and Literature Review," the second chapter, provides a brief overview of the literature on relevant research for community identification techniques, such as modularity-based, hierarchical algorithms, and spectral and partitional clustering approaches.

techniques, the model-based dynamic algorithm approach, and the supported graph feature-based techniques. It also includes current research on the topic of community identification on various social media platforms. Additionally, the provided work's issue description, study methods, and discovered research need are all discussed.

(iii) A detailed explanation of the framework for organizing communities on social media given in Chapter 3, "Semantic Relation Based Optimized Community Detecting Algorithm for Heterogeneous Network." In addition, the computational framework that is used to evaluate the effectiveness bench mark community identification algorithms on eight authentic network-based data set.

(iv) The node-centric and group-centric features on social media are thoroughly described in Chapter 4, "Single and Multi-Purpose Func Based Community Detection do Social Media." Additionally, it provides a thorough explanation of studies conducted to determine the impact of graphical features on community detection.

(v) A detailed explanation of the framework for organizing communities over is given in Chapter 5, "Multi-Purpose function base Parliamentary Optimization for Community Detect Social Media."

Additionally, it explains a mathematical and graphical method for choosing modular, dense, and resilient cliques as seed community a use influence maximization to cluster related nodes to get the ideal community structure. Additionally, the thorough experiment that was conducted work that is being given is also offered.

(vi) The "Conclusions" chapter of Chapter 6 offers a succinct summary of all this research done to develop an Influence Maximization framework Detection. The constraints that exist now and the scope of future effort for organizing social internet communities are also emphasized.

Chapter 2
Literature Review and Background

This literature1 review aim provide insight into the dynamic and varied nature or social media for community 6detection methods across multimode , multidimensional social platforms. Additionally, the aim of the literature review is Discovering fragmented communities social media requires a graphical and with clustering method.

The purpose the literature review was to identify social media feature extraction methods and graphical tools for community structuring social media. By examining graphical ands social theory base community discovery approaches, this literature review concludes by pointing out a recent research need.

2.1 Semantic Relation Basic Community Detection

Graph structures representing persons and edges reflecting relationships between people may be used to simulate social network nodes. Networks are viewed as graphs in topology-based community discovery. But the fundamental components of the issue Depending on the techniques used, different groups of the charts are created. This is the reason it is important to specify the structure no the communities that will be developed and become a part of graph initially. Making sure that the community's connection density within the community too be formed is higher than the connection density outside the communities is must common strategy. Another technique is to cluster nearby nodes by figuring out how far apart two nodes are base on predetermined similarity standards.

Some traditional data clustering methods are utilized as a result. Unfortunately, NP-hard issues characterize a large no of data clustering tasks. Because on this, several approach algorithms are used go design solutions that attempt to approximate the optimal answer. with minimal intricacy.

Grouping techniques such graph segmentation, spectral clustering, Girvan Newman divisive algorithms, modularity-based optimization algorithms, and statistics-based techniques are used go the ensemble detection issue. One go the first techniques for graph segmentation is KernighanL in algorithm [1], which finds the community yb partitioning the graph on it two equal sections at random and then sequentially changing the nnode pairs each portion. The bisection of spectral functions

based on the limits of the Fiieedler vector coordinates, approach is one of the most used techniques [2].The Girvan-Newman [3] method, which divides based on the middle link value (edge betweenness), is an illustration of a discriminating algorithm.

Numerous heuristic optimization techniques have also been devised, as the ensemble detect issue in complex networks a NP-hard problem. These techniques make use of the modularity function as the goal function tot determine the optimal way to partition the network into communities. For instance, in [4], communities are separated or amalgamated with creasing modularity for global optimization, while a simulated backgammon strategy has boon created for local optimization where a node is assigned from one community to another.

Furthermore, a variety of heuristic techniques yare employed for community discovery problems [6, 7], including the tabu search algorithm [5]. The modularity optimization algorithm is an additional technique that is commonly employed in research and is one of the modularity-based techniques. A stopping condition for a greedy hierarchical clustering technique is the notion of modularity [8].

Another method for detecting communities is to group social objects into the network by analyzing their content. The interests of people have been compiled into subject models, which group people into communities based on shared interests. The Latent Dtririchlet Assignment (LDA) is the most well-known subject model [9].

For social networks, topic models are created and understood. The models CUT [110], CART [11], and CT [122] are used to organize network users based on shared interests. The approaches to the community detection problem that rely on the network's structure and the material that its members share are relatively recent research. Taking into account the aforementioned techniques, this work presents an enhanced modularity An optimization approach has been researched to raise the modularity value. It makes advantage of the network's connection structure and the sharing commonalities among its members.

2.2 Single Multi-Purpose Function Basic Community

Detection

Today, complex network analysis is applied n0 many fields, including the study of individual land social group structures and behaviors inner the media (separation, and clustering, connection determination), online advertising and electronic commerce (customer The generation of profiles and trending analysis, customized advertising and offerings, physical structure analysis (transportation,, infrastructure), and analysis sizable data setting (media monitoring, scholarly publication analysis, genetic research) are some examples of these activities [13].

Finding communities and in networks this the most pressing problem in network research today. Identifying their this communities in on networks is used in a variety of domains, including engineering, physics, chemistry, social sciences, and biology. For instance, functional units on of proteins can be identified or their roles might be anticipated thanks to the discovery on of biological vertices communities [14]. Given the vaccine interventions for infectious illnesses in linked networks and knowledge of viral dissemination in social networks, community structure is a crucial topological aspect in sociology [15].

The majority of methods employed in the community detection are based of the division of relationships between groups. The overlapping condition known as the possibility may of nodes belonging overlapping networks is the largest issue that arises in realty network architectures. several groupings. Nonetheless, because of the intricacy of the procedures, a lot of algorithms often include nodes ways in a group while disregarding overlap [16]. Accurate information on complicated networks' structures cannot be obtained using this grouping [17]. For the purpose the of identifying overlapping communities of in intricate networks, several methods exist. The most popularity algorithm is CPM. CPM isn't adaptable enough for actual networks, though. Significant clicks are detected by CPM in crowded networks. although not in a sparse network. As a result, CPM is highly dependent on network capacity. Genetic Algorithm is used by GA-Net plus [18] adopt overlapping to communities.

process creates an line chart from a node chart. The line chart's nodes display the node chart's edges, and the node chart's edges display the edges' neighborhood relationships [19]. Those line charter is then shown as a overview of the evolutionary process, and in order to achieve fit, it is transformed into a node chart at each stage [20].

Other popular research for community identification include a technique for fast identifying overlapping ensembles [23], detection of lapping ensembles in bit networks [22], and detection are of network communities [21, 22].

Algorithms for optimization are another technique for finding communities in social networks. Finding the optimal solution to problem is their process of optimization. Heuristic optimization uniform algorithms, which widely employed in daily life, form the basis for meta-heuristic optimization algorithms, a type of decision mechanism [24]. Making decisions at crossroads and acting on direction alone when travel from one two location too another, for instance, is example of intuitive technique. Meta-heuristic algorithms are the framework that determines which approaches to use when three heuristic algorithms are beneficial for a problem from various perspectives.

2.3 Social Theories Basic Unsupervised Community

Detection

A prototype-based approach for overlapping community discovery, known as the Median variation of Evidential C-means (MECM), is presented by Karimi et al. [14]. MECM examines the produced creedal image and loosens the constraints of a metrical space embedding four the items. graph divisions to aid in comprehending the graph architecture. To prevent issues, MECM operates on a single center communist network and disregards "multi-center." The Link-Block-Topic model is presented by Chen et al. [16] for the identification of overlapping communities. It is based on this Link (LFT) model and is independent of context sampling, community count, and LFT model. The semantic relationship is established by this clustering approach.

weight (SLW), which is determined by the LFT analysis and uses context sampling and community count to separately divide the SSN into clustering units.

An technique for clustering based on entropy centrality was introduced by Fang et al. [15]. outlines the benefits of a discrete, occur random Markovian transpose process-based variation the entropy over the path-based network that was first proposed. The clustering method found communities within previously published literature and calculated a weighted dataset are for character co-occurrences in the "Les Misérables" text.

In order to examine people's actions and the connections between them, Singh et al. [17] use user interaction and profile data from several social networking sites to identify people's underlying community structure.

A parameter-free community discovery approach (K-rank-D) is presented by Liu et al. [18].

This approach determines the number communities and starting seeds from the making graph after using the PageRank centrality algorithm to assess a node's significance.

However, K-rank-D finds it extremely difficult to extract the ideal amount of communities.

A fuzzy duo-centric community detection model is presented by Shen et al. [19] in order to identify overlapping duo-centric communities in complicated networks. A de centric community is constructed around two core nodes that are sufficiently connected to one another to form the community's center. The membership values of the network's nodes are type-2 fuzzy integers that represent upper and lover membership values, respectively, indicating the degree belonging to central nodes. Interval type fuzzy membership values done characterize how nodes this shared and generate overlapping communities this communities without clear borders.

In order too define the structure off real-world great-scale networks, He et al. [22] develop a community forestry model for disjoint community discovery based on social and biological features. on order to determine the strength both similarity of edges and vertices, the community forest decompose model uses backbone degree. An algorithm based on bone degree and expansion ability is then built to identify communities from actual social networks.

2.4 Research Gap

The researcher has been worked hard lately to enhance community detection's quality and scalability. Aside from that, the variety of user-generated material on social media continues to close the research gap in the ways listed below.

(A) Handling Noise: Over a network, redundant and complementary information about network elements behave as noise. Correlations between many types of objects are displayed in a multimode network, such as Users with same interests are this likely to have comparable tags. Complementary information is present in multidimensional huge network at several levels. For instance, users may remark on one other's images even though they don't often exchange emails. Heterogeneity has recently been shown by researchers to help decrease noise [41].

(B) Communities into Multi Mode Networks: In particular, multimode community identification has great potential to shed light on networks that growing more intricate as result of social media's expansion.

groups for every mode. When it coming to displaying complicated social media and other is communication data, multimode networks are quite helpful.

pure network model, which makes analysis easier, can elegantly meet the rising data needs of more complex social and technological interactions that take place online. It makes sense that this representation can be useful in approaches used in disciplines other than social network research. As the size of the data amount sets for identifying communities in to many communities increases, more complex new algorithms are needed required to extract insights from the data.

(C) Multidimensional Network Communities: Multiple connections exist in multi to dimensional networks between two nodes, indicating different interactions (dimensions) betweenness them. Different types connections (people could be connected because they friends, coworkers, teammates, and so on) or different issue values of a single specific relation (between two authors could occur over a number of years, example) are two ways that multidimensionality in networks can be expressed. Users' social media engaged at many social media platforms, resulting in a shared community drawing in multiform networks (D). multidimensional network shares a hidden community structure, and a leader member sharing comparable passions. The primary objective is to determine the common community structure through the integration of various characteristics of network information.

2.5 Problem Statement

Social media community discovery is a crucial and interesting task. According the literature, the current methods classify end users a certain community based on their user-generated data and profile. Still, The power and modularity of these communities are compromised by a smaller number of passive, like-minded users. The current methods accommodate passive users who don't actively engage in discussions but have similar opinions on politics, goods, local and global issues, etc.

social theory-based explicit community find approach that encapsulates passive like-minded people over the world community by using graph theory and social dis joint theories for the influence maximization the social world network is required to extract influential community from social media platforms. Since the most powerful users are recognized to improve the flow on influence within the community, scalability in large networks another community identification issue that is taken into consideration.

2.6 Research Methodology

The study methodology that follows was created to address the graph view point social media in order to determine the trade-off betweenness the density of connections and the similarity of nodes' qualities using influence maximization.

(a) A thorough literature review has been conducted in order to get and comprehend the varied nature of internet media data.

- For multimodal and multidimensional social world platform community detection.

- To concentrate on the problem of social world mining in order to extract graph-based information for

user-generated content and personal data social media platforms. order to examine how graphical feature affects community detection performance, a computing framework is constructed that effectively For clique structure, extract the group- and node-centric features as seed

community. - Compares features that are group-centric and node-centric using actual network huge data sets.

- Provide intersecting information on the strengths both weaknesses of group-centric and node-centric features for community discovery.

(c) A social theory-based influence maximization ideal framework (IMCD), which incorporates social theory—that is, social balance, quality, status, and correlated theory—has been sculpted in this order to maximize the modularity, social and mutual information of the internet community.

- Use a similarity index based on Jaccard coefficient to identify identical nodes.

- Use influence propagation to locate the passive node.

(d) The performance of the created framework is examined and contrasted using standard evaluation parameters.

(e) The impact maximization framework ,graph for community discovery improved the social atomic clustering low performance based on technique that was described.

Chapter 3

Heterogeneous Network-Specific Semantic Relation Based Optimized Community Detection Algorithm

3.1 Introduction

One kind of the issues in social networks has received painful the most attention lately is community detection. Numerous disciplines, including sociology, marketing, and security, depend on the analysis of interpersonal connections, hierarchies, and shared interests [44, 45]. Link analysis techniques typically used case in existing research on the community recognition challenge [1-3,46]. Because the community is characterized as a collection of entities that share certain features, the aforementioned similarity metric differs depending on the network.

Conventional approaches often divide the network into communities while taking use of its topological characteristics. The optimal network fragmentation into communities is the primary criterion upon which the proposed approaches are predicated. Because of this, the merits of the various strategies in internet are compared using the modularity measure [8]. To raise the modularity value, many optimization-based approaches are being developed. Using content analysis to classify people based on shared interests is another method used in community discovery [9–12]. People that work use social networks exchange several types of relevant and actionable data, including emails, blogs, and media shares. People with comparable data are assumed to have similar interests. Using clustering techniques, the network's community may be identified based on the qualities gleaned from the shares. Examining the two strategies, it is evident that the techniques that make advantage of the Users' interests are ignored by the network's topological properties. On the other hand, approaches that concentrate on content analysis disregard the network's topological structure [47]. Two-stage community identification techniques are being developed as a solution to this issue [4]. Following the network's dissection based on content analysis, two-stage methodologies apply link analysis to each community that is obtained. By combining the network's content analysis and link analysis, the study goal of this chapter is to try to raise the network's modularity value.

As a result, the network of joint was constructed using the similarity values of two between individuals' shares as indirect linkages. The developed technique makes advantage of indirect connections of some with network structure and a greedy also some time hierarchical clustering algorithm. The modularity-based modularity optimization [8] technique has been reformulated in light of the semantic commonalities among the network's participants. The nodes that are topologically nearest to one another are grouped via the modularity optimization technique. However, our method are makes sure to that the nodes are in that are most similar to one another in terms some of both topology and semantics are prioritized.

3.2 Parliamentary Optimizations Algorithm (POA))

The first phase in the optimization process of the Parliamentary Optimization process is to create the starting population of people. The persons that are produced are regarded as parliamentarians. The populace is split into political groups the next stage. groupings (parties), and candidate with the best suitability for the stipulated number of person groups is taken into consideration. The ingroup composition begins after this stage. Several individuals with this highest qualifications are chosen as the final person for any group at conclusion of the in-group competition process. The chosen candidates fight against those from other groups in the following round. The group's principal and potential members are crucial in establishing the group's overall influence. Group competition begins after the intra-group competition phase. Parliamentary political parties with one to another to support their nominees. order to improve their odds of success, strong groups occasionally come together and form one cohesive entity. Algorithm 1 illustrates the POA process phases.

Algorithm 1: Parliamentary Optimization Algorithm (POA) step-by-step explanation:

1: Get going

2: The starting populace is established.

3. (a) There M groups made up L people the population.

(b) The most physically fit person is chosen the group's candidate.

(b) Intergroup rivalry

4: (a) Prominent members approach each group's candidate members.

(b) Appointing new candidates.

(c) Determine each group's power.

5: Group competition

1. The most powerful group is identified, and PM probability is applied to these groupings.

2. Pd, the weakest group, will probably be eliminated.

6: Step 3 repeated if the end condition are not satisfied.

7: The optimal choice is thought to be the answer to the optimization issue.

8: Come to an end.

3.3 Modularity Optimization Algorithm:

procedure increases the modularity on value by combining pairs of nodes, which is an hierarchical clustering method technique. The number ensembles is indicated by the union with the is highest modularity score. The circle is defined by the algorithm as a graph, and it determines

which node pairings, when combined with the network's connection topology, will has the highest modular value. The procedure is described in Algorithm . graph with the form G = (V, E), here V represents nodes , E represents edges, is referred as a social network. The social network's users are represented the set of nodes V, and their relationships are represented by the collection edges E. The modularity valuation Q of an unweighted, undirected graphics G with m edges may be found using equation (1).

collection of edges E represents the connections between users. Equation (1) gives the modularity value Q for an unweighted and undirected graph G containing m number of edges.

$$Q = \frac{1}{2m} \sum_{uv} \left[A_{uv} - \frac{d_u d_v}{2m} \right] \delta \left[r(u)\, r(v) \right] \qquad (3.1)$$

The node u , v's inclusion in this ensemble r is indicated by the values of r(u) and r(v). When i = j, the function δ yields δ[i, j] = 1, and when i = 0 yields 0, respectively. The degrees of the and v nodes' nodes are denoted by du and dv, respectively. Regarding the neighborhood In matrix A, the node you u and ,v are near the situation when Auv = 1. A direct link is one this exists between 2two nodes. The modularity optimization process creates as many classes as there are nodes with ever node being designated as a distinct class. Combining the grouping that raise the partiality modular value (ΔQ) each step yields significant groupings. The effect off a merger modularity is shown by the ΔQ value. Modularity value (Q) increases when two communities high ΔQ values are combined. Equation (2) gives the value ΔQ.The side fractions is Pj eij with the orgin is equal ai.

$$\Delta Q = 2 (ei\, j - aiaj)$$

The algorithm Enter: V,E Publication: Communities

1 Make k classes, then give a node to every class.

2. Determine the value initial modularity. Q. 3: Sets to 1–4 and repeat s = k).

5 J and I. Determine ΔQij values the class.

Six Put the courses with the greatest ΔQ together.

Give up.

3.4 Semantic Relation Basic Optimization Algorithm

This section provides an explanation of the Parliamentary is optimization technique and modularity that was established via the use is of semantic similarity. Priority grouping is of nodes this are topologically semantically nearest to one another is provided by the established method. The degree of resemblance between documents exchanged by persons on a network (such as emails, blogs, etc.) is known semantic closeness between members. Every member of this network is represented by word distributions. As an indirect link, the similarity (Cuv) .

Let the collection z documents that belong to u be represented by du = du1, du2,..., duz. To determine each person's word distribution inside network, u. Every document associated with the node regarded a single document.Du= w1,w2,...,wk u. denotes k sets words that are part of node. Equation (3) use the Cosine similar approach [48] to determine indirect connection value, or semantic similarity, between u and v.

between nodes u and v is calculated using the Cosine similarity method [48](Equation (3)).

$$C\left(d_u, d_v\right) = \frac{d_u * d_v}{|d_u| * |d_v|} \tag{3.3}$$

Pattern detection and vector representation of words are required to detect best the similarities and differences between the word distributions of the members. In the

The most effective way to identify this similarities and of differences between members' word distributions is using pattern recognition and vector are representation of words. Meaningless terms from the members' documents were taken out for the study. Every every term that was utilized in the network's root was found. The words were weighted based the members' frequency of usage using TFIDF algorithm [49].

n ∗ n C size similarity matrix produced for n-node networker using the cosinenity similarity approach. Cij I in conjunction with J. It illustrates how similar the classes' semantics are to one another. The created approach starts with an n−n dimensional C matrix and ΔQ. Multiplying the ΔQij and Cij values yields the combined strength (ΔQCij) of i and j classes. The greatest ΔQC value of two categories merged. Following each merging operation, the C is modified, as demonstrated by algorithms 3 and 4. The average of two types' similarities is used get the similarity values the combined class. ΔQCij is given the lower similarity value the Cosine matrix if the two members' calculated sharing similarity, Cij, is originally equal to 0.

3.5 Analysis of the Experimental Setup and Results

Performance evaluation have been done across six distinct graphical social data sets, including Word adjacencies, Zach Karate Club [50], Dolphinne social media, and others, to determine the community discovery algorithms. books on US politics, American college football assessment parameter modularity. SR-MOA Algorithm 3 The algorithm Enter: V, E, Cn Publication:

1 Make k classes give each class a node assignment.

2 Determine the value initial modularity. Q. 3: Sets to 1–4 and repeat (s = k).

5 J and I. Determine the values of ΔQij for class 6. Determine ΔQCij = ΔQij ∗ Cij.

7. Class combinations with the greatest ΔQC 8 Update are C matrix full Stop are combined.

SR-POA Algorithm 4 The algorithm Enter: V, E, Cn Publication: Communities

1. Get going

2. Establish k classes provide a node to every class.

3. Determine the value initial modularity, Q.

4. The item is set to 1.

5. Continue (until s = k).

6. The starting populace is established.

(a) The population is split up into M groups, each with L people.

(b) The most physically fit person is chosen the group's candidate.

(b) Intergroup rivalry

7. (A) well-known members approach each group's prospective members.

(b) Appointing new candidates.

(c) Determine each group's power.

8. Group competition

(a) These groups joined using Pm probability, and the most attracting influential group is identified.

(b) It is probable that the weakest Pd will be removed.

9. Step 3 repeated if termination condition not satisfied.

10. The optimal choice is thought to be the answer to the optimization issue.

11. Come to an end. In order to extract community structure from a network, modularity—a network structural measurement—evaluates the strength subgraphs, or groups, clusters, or communities in network [53]. Groups nodes with more modularity in network are often dense in each This results in the following communities' presence in given network:

Table 3.1: Comparative Analysis of Impact of Semantic Relation on Modularity

Classification Technique	Modularity					
	ZKC	ACF	DCN	BUP	LM	WA
MOA	0.4512	0.5568	0.5684	0.6575	0.5423	0.3956
POA	0.6045	0.6856	0.6423	0.6953	0.6845	0.4956
SR-MOA	0.5245	0.6435	0.6462	0.7512	0.6075	0.5056
SR-POA	0.7014	0.7856	0.7126	0.7856	0.7456	0.6586

Modularity is network structural measurement that evaluate the strength of sub graph (groups, clusters or communities) in network for extracting community structure [53]. In a network, group of node having higher modularity are relatively dense each other and leads to the appearance of communities in a given network as :

$$M = \frac{1}{2|E|}\sum_{xy}[e_{xy} - \frac{w_x w_y}{2|E|}]\delta(c_x, c_y) = \sum_{i=1}^{n}(f_{ii} - f_i'^2) \tag{3.4}$$

Where e_{xy} represents the edge from node x to node y, W_x represent the summation of the weights of the edges linked to node x, c_x is the belonging community structure of node x,(c_x , c_y) is a probabilistic function that equals to 1 if both the respective node x and y belong to same community structure, otherwise 0. fii represent the edge in community i and F_i' is the belonging probability of random edge to community i that attached to vertices in community i.

Whereas, Normalized mutual information is a normalization of intra-community mutual information score to scale the similarity between intra community node as:

$$nmi(x,c) = \begin{cases} 0 & node\ are\ totally\ dissimilar \\ 1 & node\ are\ totally\ similar \end{cases} \tag{3.5}$$

and mutual information is calculated as

$$nmi(x,c) = \frac{2 * i(x,c_i)}{e(x) + e(c)} \tag{3.6}$$

where the label is denoted by x, the community shape by c, the entropy by e, and the information for ci for class x is represented by i(x;c).

Assessment of community discovery algorithm's performance both with and out semantic connection are displayed Modularity and of course Normalized .

Table 3.2: Comparative Analysis of Impact of Semantic Relation on Normalized Mutual Information

Classification Technique	Normalized Mutual Information					
	ZKC	ACF	DCN	BUP	LM	WA
MOA	0.7245	0.6845	0.6256	0.4567	0.3845	0.4125
POA	0.7445	0.7056	0.6645	0.5124	0.4756	0.4856
SR-MOA	0.8178	0.7456	0.6856	0.5042	0.4123	0.4986
SR-POA	0.8456	0.7635	0.7456	0.6845	0.5945	0.5896

information, respectively. Both the evaluation parameter is significantly improved after incorporating social theories with community detection algorithm.

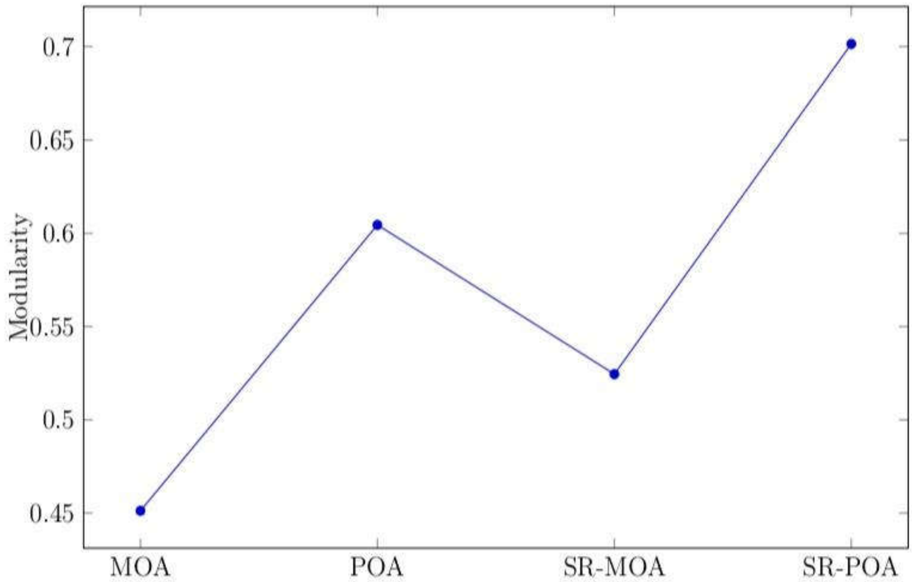


Figure 3.1: Modularity of Community Detection Over ZKC Data Set

data, correspondingly. Following the integration of social with the detection algorithm, both assessment parameters show a considerable improvement.

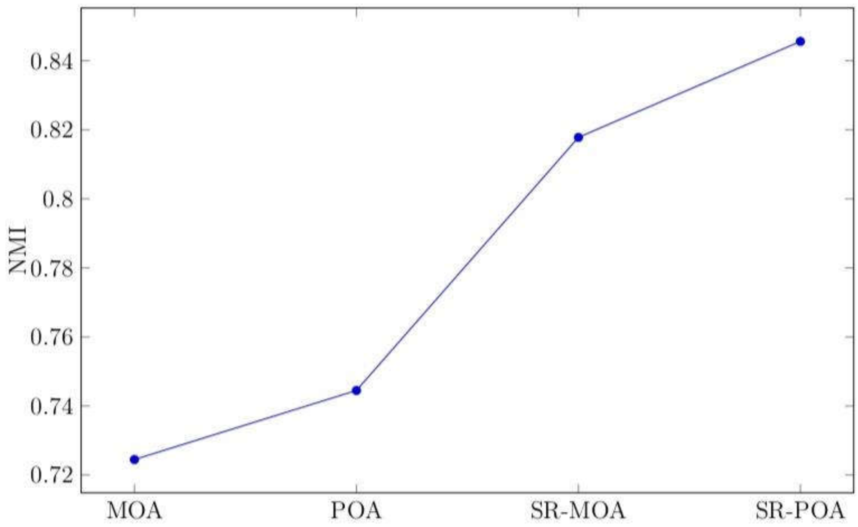


Figure 3.2: Normalized Mutual Information of Community Detection Over ZKC Data Set

In contrast, as figure 3. and 3.4 illustrate, the community discovery algorithm MOP and POA acquire 55.68%, 67.56% modularity and 63.45%, 70.56% NMI over the AFC data , respectively. However, with the incorporation of semantic relations (SR-POA and SR-MOA), The AFC data yielded much higher modularity NMI values, acquiring 64.35 percent and 78.56% of modularity and 73.56% and 76.35% of NMI, respectively. In contrast, , community detection MOP and POA gain 57.84%, 64.23%, and 62.56%, 66.45% NMI, respectively, over the DCN data set. On other hand, modularity and information are greatly enhanced and gained 64.62%, 72.26% modularity 68.56%, 73.56% NMI DCN set, respectively, after including semantic relation, i.e., SR-POA and SR-MOA. The highest NMI informed is achieved by the SBA HSA algorithms, but the SEOA algorithm in modularity. On other hand, as fig 3.7 and 3. illustrate, the community discovery algorithm MOP and POA acquire 65.75%, 68.53% modularity and 47.67%, 51.24% NMI over the BUP data set, respectively. However, the modularity and NMI information are much enhanced and obtained 75.12% by integrating semantic relation, i.e. SR-POA and SR-MOA,

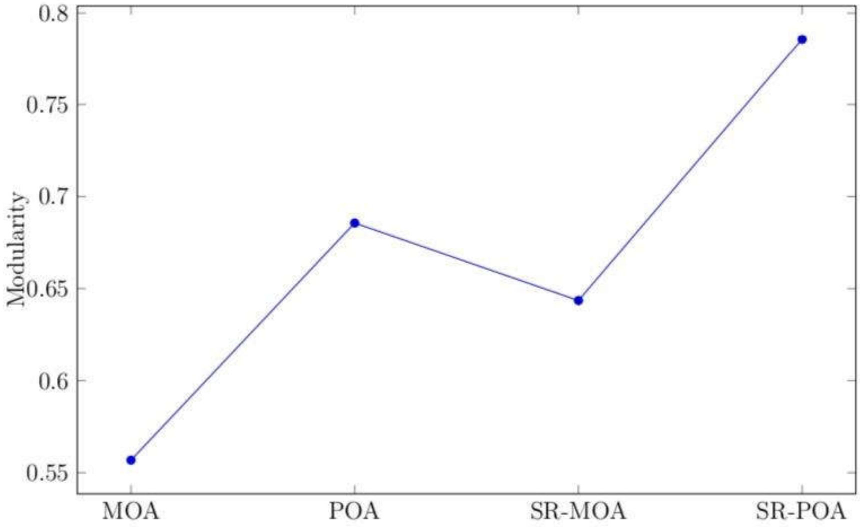


Figure 3.3: Modularity of Community Detection Over AFC Data Set

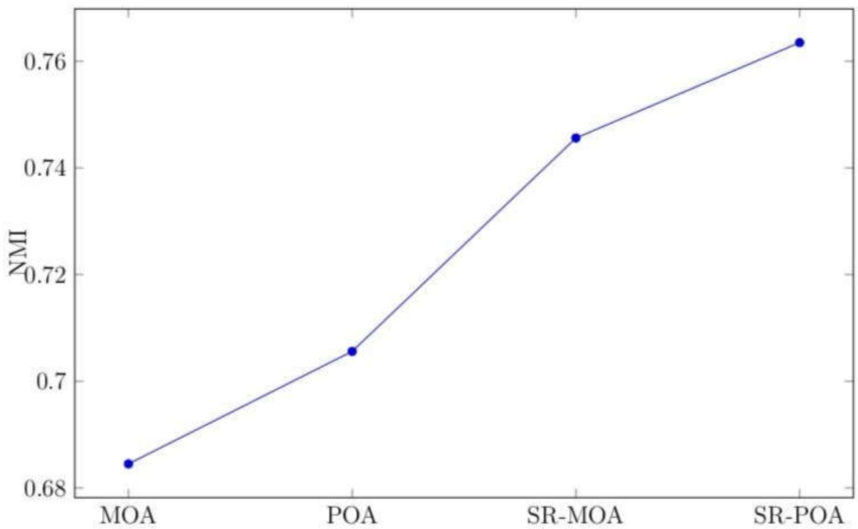


Figure 3.4: Normalized Mutual Information of Community Detection Over AFC Data Set

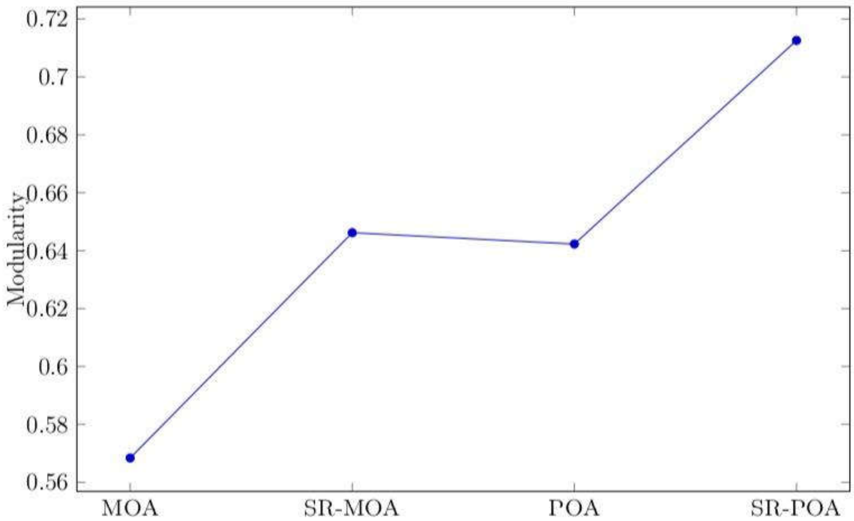


Figure 3.5: Modularity of Community Detection Over DCN Data Set

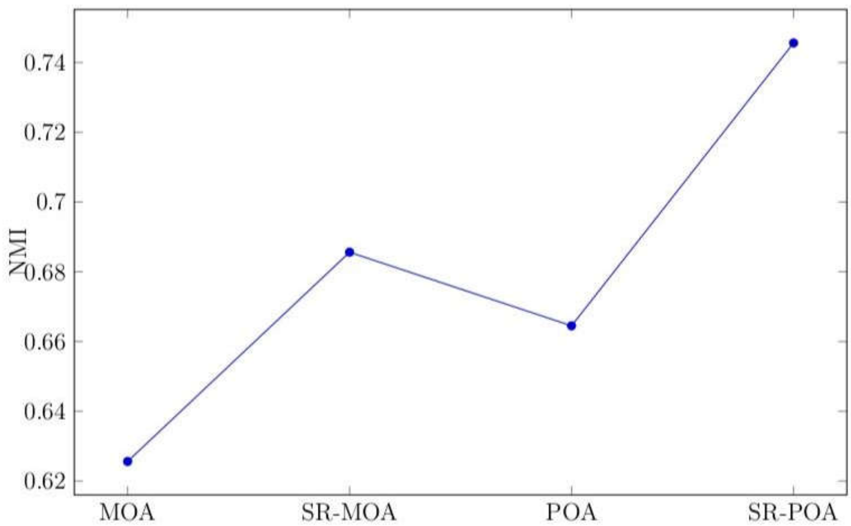


Figure 3.6: Normalized Mutual Information of Community Detection Over DCN Data Set

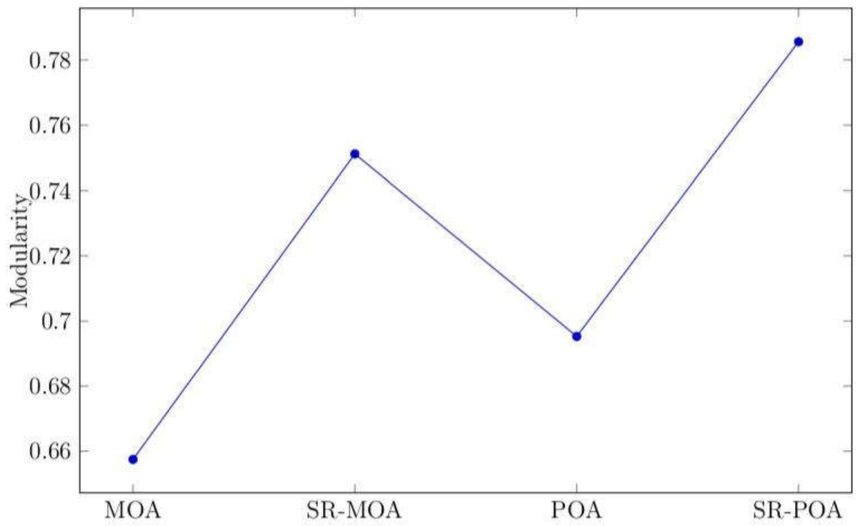
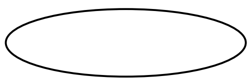


Figure 3.7: Modularity of Community Detection Over BUP Data Set

78.56% modularity and, across the BUP data set, 50.42%, 68.45% NMI, respectively. On the other hand, as figures 3.9 and 3.10 illustrate, the community detection algorithms MOP and POA gain 54.23%, 68.45%, and 38.45%, 47.56% modularity and NMI, respectively, over the LM data set. However, the modularity and NMI information are considerably enlarged and obtained 60.75%, 74.56% modularity and 41.23%, 59.45% NMI over LM data set, respectively, after including semantic connection, i.e. SR-POA and SR-MOA. On the other hand, as figures 3.11 and 3.12 illustrate, the over WA data set, community detection technique MOP, and POA gain 39.56%, 49.56% modularity and 41.25%, 48.56% NMI, respectively. On the other hand, modularity and NMI information are greatly enhanced and gained 50.56%, 65.86% modularity and 49.86%, 58.96% NMI over WA data set, respectively, after including semantic connection, i.e., SR-POA and SR-MOA.



Community circle

Fig:3.9

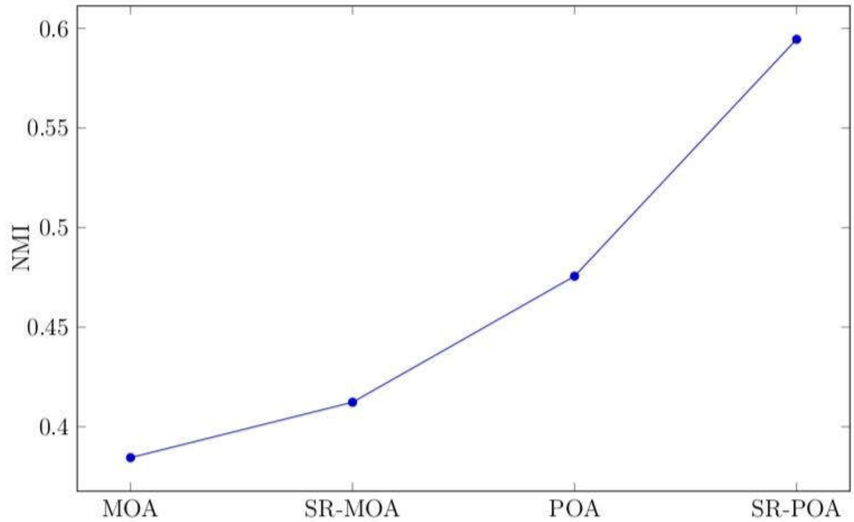


Figure 3.10: Normalized Mutual Information of Community Detection Over LM Data Set

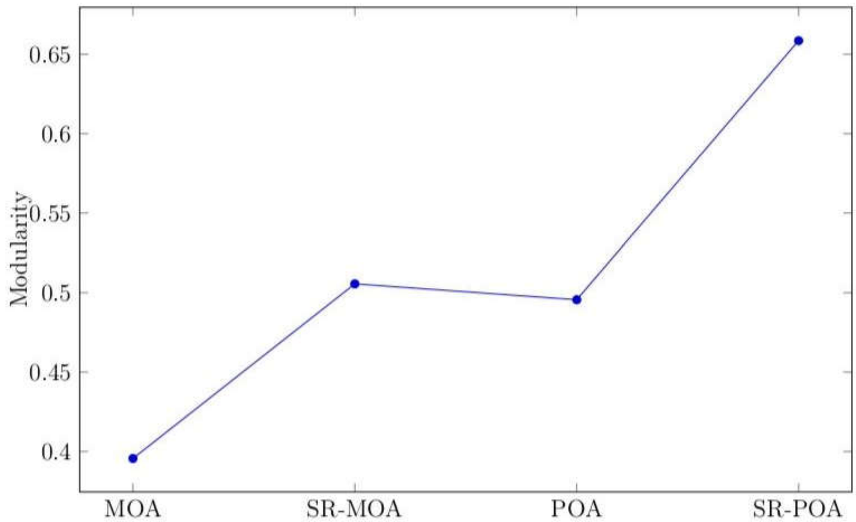


Figure 3.11: Modularity of Community Detection Over WA Data Set

Figure 3.9: Modularity of Community Detection Over LM Data Set

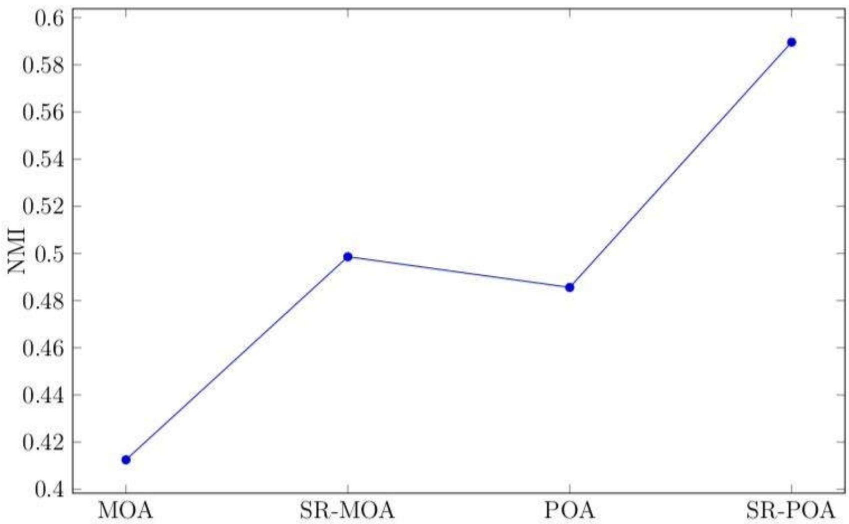


Figure 3.12: Normalized Mutual Information of Community Detection Over WA Data Set

With semantic relations, the POA and MOA algorithms perform much better and obtain higher performance rates over the more densely packed ACF and ZKC networks, whereas the performance of the latter is comparatively lower over the less densely packed WA data set.

to the topic of community recognition using network link structure analysis. The produced communities in this instance solely take into account the network's topological features; the shared documents between network users are disregarded. In this study, a content analysis of the network was conducted in conjunction with the topological structure of the network to leverage semantic commonalities among individuals. The findings demonstrate that improving semantic similarity between people and link structure improves the community detection problem's performance. Future research examine how alternative optimization strategies, utilizing the semantic similarity approach, affect modularity performance.

Chapter 4
Single and Multi-Purpose Function Based Community Detection over Social Media

4.1 Introduction

The 21st century has seen a tremendous advancement in communication and technology, making it essential for individuals to have access to and utilize information efficiently. This has made information a necessary element of daily life. The quickest and most The internet is without a doubt the most practical means to get information in the modern world. In addition to being a network that links millions of computers globally, the Internet is an ecosystem that links millions of individuals and thousands of social groupings, and it is always expanding. One of the most widely used internet apps, social media is quickly rising to prominence as one of the most crucial communication tools available today. The rate at which people social media rises in tandem with the frequency at which people the Internet. It's predicted that media will eventually supply nearly all the internet's needs. Social media apps aim to meet almost everyone's needs by covering a wide range of subjects, including games, finding information, and searching, in addition to conversation. People won't need another tool if they can find nearly whatever they're looking for on social media. Researchers may now access and analyze data from vast networks, such as social media, thanks to advancements in computer technology and network analysis.

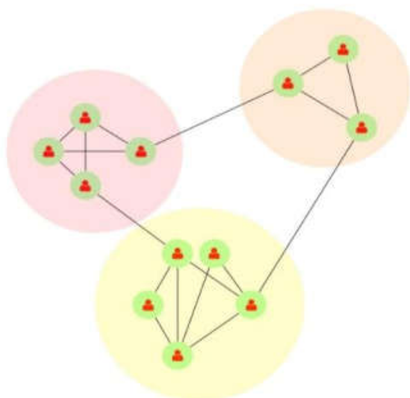


Figure 4.1: Network structure consisting of three communities

The ability classify nodes based on structures of groups belong to and to expose groupings is a crucial aspect of community discovery. In the event that a node-set in a society network structure has more linkages than this node-set is regarded as a community due to its external connections. Communities, which often referred as clusters or modules, are collections of nodes that operate together in networks and typically have common functionalities [54]. Figure 4.1 provides a grid that graphically depicts the communities. The majority of community detection techniques are based on the division of linkages between groups. The most common issue in real-world network architectures is what's known as the overlapping scenario, or the potential for nodes to be part of several groups. Due the intricacy of the procedures, many algorithms, however, often include nodes in a group, disregarding the overlap [16]. Accurate information on the structure of complicated networks cannot be obtained with this grouping [17].

4.2 Community on Social Networks

Networks are used depict the majority of complex networks. For instance, networks are networks in which individuals are represented as nodes and links between them by other people; the World Wide Web (WWW) is a network of linked web pages. sides. Comparably, biological networks are made up of nodes that represent biochemical molecules and borders that indicate the connections among them [55]. The majority of research in recent has been are devoted to comprehending how network topology affects system dynamics, behavior, and network organization and development. Understanding complicated network architectures also requires identifying community structures [56]. Communities in a network are node groupings with frequent linkages within groups and rare connections between groups. A other definition of the community is an organization of people communicate with other on a regular basis. Because of this, communities are essentially collections of nodes that interact similarly and have comparable qualities [3]. Communities inside network architectures provide us specific information about people's study topics, interests, patterns, and other characteristics. The network structure is not uniform in actual networks. Clusters of nodes with comparable roles and shared features are likely what make up systems that concentrate and cluster in a certain region, which we refer to as ensembles [54]. There are several concrete application areas in communities. By allocating the same servers to every client cluster, clustering web clients with comparable interests or those who are physically close to one another, for instance, improves service performance on the WWW. In online buying systems, an efficient advice system between the buyer and the vendor may be formed by identifying the community of consumers with similar interests [45].By recognizing communities, hierarchical organizations in intricate real-world networks may be planned. Communities within communities are a common feature in real networks. The human body is the perfect example of a hierarchical structure. The body is composed of organs, organ-derived tissues, and tissues composed of cells. Business firms are another instance of a hierarchical organization. Businesses with mid-level workgroups can be compared to a pyramid where the workers at the base develop to become the head of the organization. By employing merely coded information of the network topology, communities in networks are intended to be described, together with their hierarchical organization of modules. In community discovery, the most commonly applied term is the presumption that the There should be more edges inside the group than there are connections outside of it. The number of edges that join this group to the remainder of the line is known as the "cut-size" parameter, which is determined starting at this point. A modest cut-size value is expected for a decent assortment.

Another meaning of "vertex similarity" is the concept that distance between nodes on a spaceplane serves as a similarity criteria. Traditional grouping techniques frequently employ this strategy. In case nodes are unable to be positioned on a spaceplane, an adjacency matrix may be employed. Even if they are not neighbors, it may be claimed that they are similar if their neighbors are the same. Furthermore, the number of independent pathways, the shortest path, or random walk two nodes may be used to calculate the similarities between them [57].

Although the first research on identifying community structures proposed that a node could only be a part of one community, networks are made up of many interactions where nodes may be a part of multiple communities; this type of structure is known as overlap. Human interactions can include, for instance, relationships between family, friends, and coworkers. Consequently, identifying overlapping groups is a crucial problem for studying actual social networks. A network including three distinct communities is presented in Figure 4.2. Four nodes in the network are part of several communities, demonstrating how communities in networks overlap in structure. Since the community and modular structure are utilized to determine the systems' performance, they are regarded as crucial components of real-world social networks. Despite the numerous ambiguities surrounding community identification, successful and efficient community finding techniques have been established.

4.2.1 Graph Partitioning

It involves splitting the nodes into k groups of fixed sizes so that there are as few edges as possible between the groups. Nevertheless, it is not an appropriate strategy for situations when the whole number of groups existing in social network topologies is unknown beforehand.

social network examination. Iterative bi sectioning is one of its most important algorithms [58],

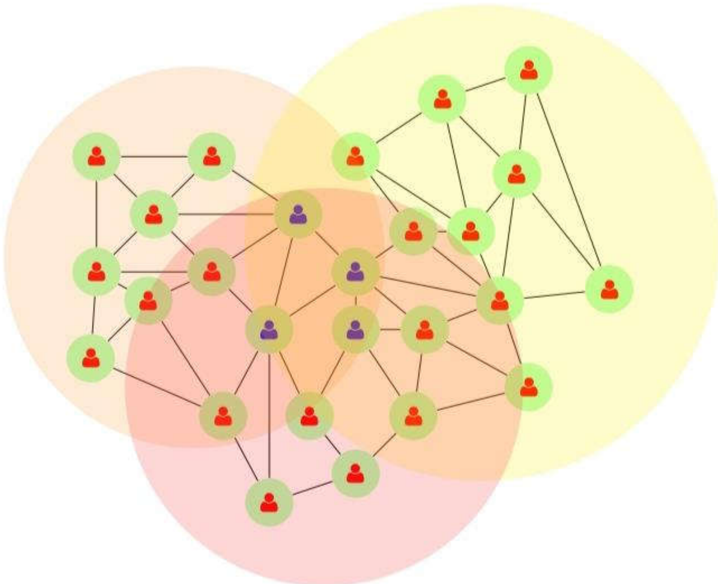


Figure 4.2: Overlapping community structure

Min-Cut Max-Flow Theorem. A dividing line for k = 2 with the fewest edges between groups is displayed in Figure 4.3.

4.2.2 Hierarchical Grouping

Groups within social networks are typically entwined in a hierarchical structure.

It is a technique based on grouping similar nodes together and removing low-affinity nodes from groups to divide them. Findings will differ based on how similar they are. To be decided criteria [54].

4.2.3 Segmentation Clustering

Each node is viewed as a point in space, and the group number k is predefined in this case.

The goal is to organize points into k groups based on how far they are from the center, based on a function that is provided. The majority Minimum k-clustering, k-central, k-median, and k-means are components that are employed [59]. The drawback is that it's necessary to know ahead of time how many groups there will be [54].

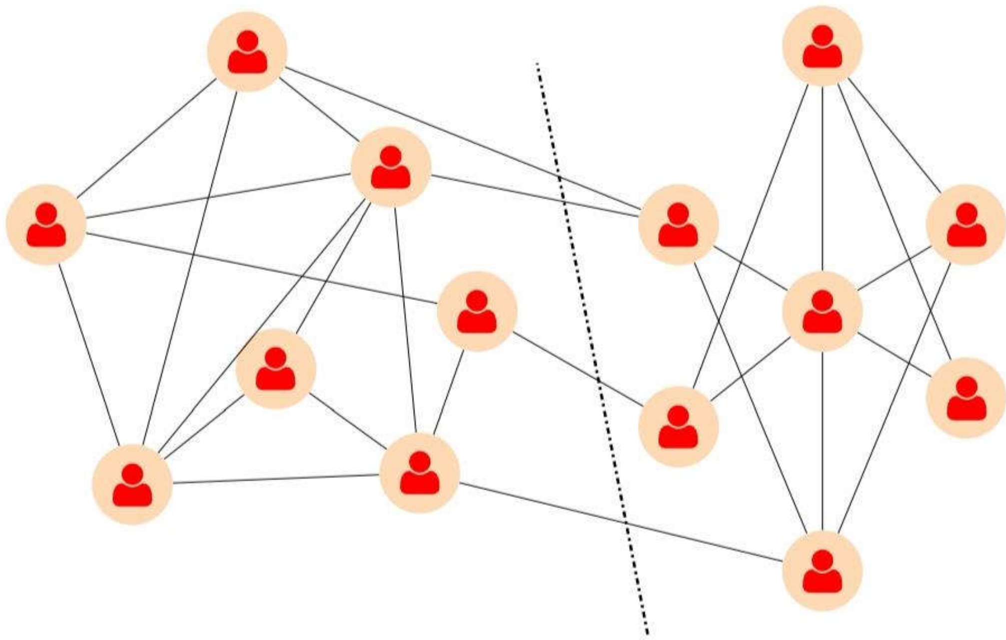


Figure 4.3: Graph Partitioning over Social Media Network

4.2.4 Spectral Clustering

Numerous strategies and procedures divide data in sets using an eigenvector, such S, or other matrices derived from it are used in spectral clustering [60]. Using this approach, groups with a similarity matrix's eigenvectors as a starting point are created. k-means function, for example [59].

The Laplace matrix is the most often utilized matrix. This method allows the number of groups in line to be determined from the eigenvector components.

4.2.5 Segmentation Algorithms

It's a technique used to identify and eliminate the edges in a graph that link groupings in order to separate and make those groups more visible. How to identify the edges that link these groupings is crucial. The Girvan-Newman algorithm is the most often used one. algorithm [61]. In this case, edges are chosen according to a criterion known as edge centrality. Every edge's centrality value is computed. Deleted are the edges with the greatest centrality value. The procedure is repeated, starting from step one and ending with the deletion of the edge with the greatest value In addition to the edge centrality criterion, other criteria that are employed include edge betweenness, random walk edge betweenness, and current flow betweenness [54].

4.2.6 Modularity Based Methods

The most popular and often applied quality function in graph analysis is modularity. A high modularity value is thought to suggest excellent groups, albeit this is not entirely confirmed [54]. If a graph is more modular than a randomly spaced line of same length and that line is seen a group structure to some extent. A high modularity rating, however, not automatically the presence of a group structure. Certain random graphs may have high modularity ratings even when they lack group structure.

Enhancing the modularity function lacks a linear time solution as it is an NP-Complete issue. Nevertheless, effective algorithms with a range of convergences have been created [62, 63]. From the set of adjustments performed on the chart, the change that maximizes the quality function is chosen. This might be an edge deletion, split, or merging.

4.2.7 Dynamic Algorithms

The random walker model is the most often utilized dynamic algorithm for community exploration. When using this strategy, the random walker remains in the community for an extended period of time if the graph's connections have a high density; according to strong communities make up the chart, which illustrates the disputed rationale [54].

4.2.8 Other Methods

In addition to the previously listed and widely used techniques, there are techniques based on statistical inference (Bayes, etc.) [64, 65], techniques that tag nodes and utilize the tag that their neighbors share the most in each iteration to divide groups in this manner [54], There are click filtering techniques [66], overlap prevention tactics [67], and multi-resolution techniques accessible.

4.3 Methodological Optimization Methods

An optimization issue is any task that involves determining unknown parameter values while adhering to specific constraints. To optimize is to optimize. Finding the best answer out of all the options for a problem under the circumstances at hand is the task at hand. In other words, while heuristic algorithms can converge, they cannot provide a precise result. This circumstance offers a resolution that is nearly perfect [68]. Heuristic algorithms are necessary for the following reasons:

It's possible that the optimization issue has a structure that makes it impossible to define the precise answer. Heuristic algorithms may be used to locate the precise solution and for learning reasons. They can also make the decision-maker's job considerably easier in terms of clarity.

* The most difficult parts of real-world issues (such as which objectives and constraints should be applied, which alternatives should be investigated, and how to gather problem data) are frequently overlooked in definitions created using mathematical formulae. More serious mistakes might result from inaccurate data employed during the model parameter determination stage than from the heuristic approach's potential for producing a suboptimal solution [68]. become increasingly powerful and well-liked. They solve search or optimization issues by using a straightforward approach. These can be explained by the following summary:

a. They provide broad approaches to solving problems that can used to the issue when various kinds of objective functions, restrictions, and decision variables are present. Strategies for finding solutions are independent of the kind of objective function, limiters, and variables that are utilized to describe the issue.

b. It is independent of delimiters, decision variable count, and kind of solution space.

c. It doesn't require highly precise mathematical models, which can be challenging to set up for the model and system's intended function and occasionally unusable due to high solution time costs.

d. They don't require a lot of computing time because they have strong processing capability.

g. They are simple to modify and adjust.

f. It produces useful outcomes for combinational and nonlinear large-scale challenges.

g. Unlike classical algorithms, a solution method for a specific issue does not need certain assumptions that may be challenging to verify in adaption.

h. It doesn't necessitate modifications to the issue of interest, just like traditional algorithms. They modify themselves to address various problems.

These benefits make meta-heuristic algorithms widely utilized in many domains, including computers, engineering, and management science, and new versions are advised.

Figure 4.4 lists the general-purpose meta-heuristic methods: bio-based (evolutionary algorithms, ant colony algorithms, bee colony algorithms, artificial immune algorithms, firefly algorithm, enzyme algorithm, sapling development algorithm, invasive weed optimization, monkey search algorithm, bacterial bait search algorithm), physics-based (multi-point heat treatment algorithm, electromagnetism algorithm, particle collision algorithm, big bang - big crash algorithm), swarm-based (bee colony optimization), social-based (multi-point taboo The group evaluates eight different methods: research algorithm, parliamentary optimization algorithm, imperialist competitor algorithm, music-based (harmony search), sports-based (league championship algorithm), chemistry-based (artificial chemical reaction optimization algorithm), mathematics-based (metaheuristic and base algorithm) and sports-based (harmony search) [70]. Hybrid techniques for integrating them also exist. Even if the literature has produced many very effective algorithms and methodologies, it is crucial to create, develop, and use new strategies under the

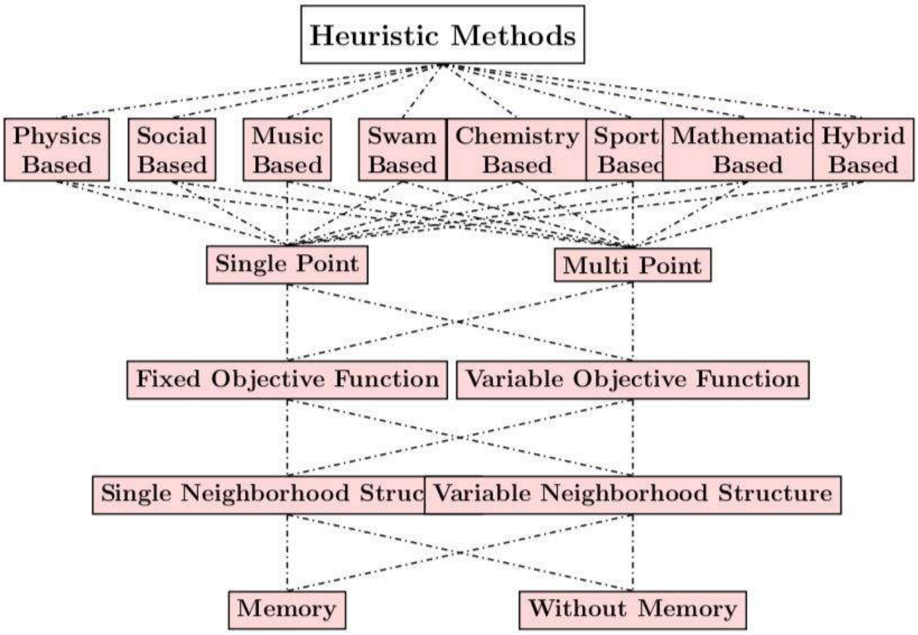


Figure 4.4: Meta-Heuristic Methods

concept that emphasizes constant progress and the pursuit of excellence in the scientific domain. Furthermore, new meta-heuristic methods are continuously being given since the algorithm that yields the best outcomes for all issues yet been designed. The Current ones are proposed to function more efficiently. Because of this understanding, researchers has effectively used novel meta-heuristic approaches and brought them to the literature in recent years.

4.3.1 Algorithms for Socially Based Metaheuristic Optimization

The literature contains a big number recently introduced social-based heuristic optimization techniques. tabu search are algorithm is most widely used and well-known of them. Others have been filed more recently [71].

4.3.2 Imperialist Competitor Algorithm

The Imperialist Competitor process (ICA) starts the process by generating an initial population, just as analogous evolutionary algorithms. a couple of the leading nations in Bhopal, Maharashtra, the individuals who remain after the initial population is selected to be imperialists become imperialist colonies. The imperialist states divide up all of the specified territory. Following their distribution among the imperialist governments, the colonies start to drift in direction the relevant imperialists. Empires are only as strong as their imperialists and the areas they have been granted. The algorithmic process is still under progress, with the race between the imperialists having begun. Having failed to get stronger or achieve success, the imperialist will be wiped out from the race. Strong empires gain power during the race, while weak kingdoms lose ground and eventually collapse. The race goes on until there is only one empire left, at which point other nations become colonies of the one that is still standing thanks to the algorithm. Territories and imperialists will hold equal status and influence in the utopian society that emerges at the conclusion of the race [72]. The algorithm's flow chart is displayed in Figure 4.5.

4.3.3 Educating Algorithms for Learning-Based Optimization

The (TLBO) method is another newly created meta-heuristic optimization method [73]. The TLBO algorithm operates based on how a teacher affects the pupils in a classroom. The algorithm explains the capacity instructors and students at a school teach and learn. Two fundamental elements of this algorithm are the teacher and the pupil [74].

The population in the algorithm is the group of students, and the many subjects that are taught to them are thought of as additional design factors for the optimization issue. The outcome of student is most comparable the optimization problem's fitness value. For entire population, the teacher is thought to be the best option. The optimal solution is represented by the best value of the fitness function, and the terms used as design variables are displayed as the parameter included in the fitness function of the given optimization issue. The two scenarios that make up the TLBO algorithm's functioning process are the teaching process and the learning process [75]. Within the context of the teaching process, the teacher is widely seen as the key figure who imparts information to the students. Students' responses reveal the quality the teacher. It has shown that where pupils have excellent professors, both their circumstances and grades improve. Thus, the Instructional Process.

4.3.4 Social-Emotional Optimization Algorithm

A novel social-based optimization method that mimics human behavior is called the Social-Emotional Optimization Algorithm (SEOA) [77]. The human community is linked to the term "social." The community's residents work to elevate their social standing. Every individual in SEOA is a virtual person. People make behavioral decisions based the corresponding emotional index at each stage [78]. There three categories the emotional index: low, medium, and high. The emotional is used to determine which behavior is chosen. Depending on the intended behavior is accurate, behavior. The person's emotional index rises if this decision raises the society status value. If not, the social value declines as the emotional indexing decrease [79].

Algorithm 1 provides the stages in the SEOA procedure.

Algorithm The Social-Emotional Optimization Algorithm is explained in five steps:

1: Get going

2. Every person is created one after the other, and the problem space is randomly assigned to each one of their starting places.

3. The goal function is used compute each person's fitness value.

4. j. Based the person's emotional index; their behavioral motions are ascertained.

5: The whole population's location is updated.

6: The emotional quotient is calculated.

7: best option is approved if termination requirement is satisfied. Step 2 returned if condition not satisfied.

8: Come to an end.

4.3.5 Brainstorming Optimization

In organizations that are generally acknowledged, brainstorming is a common method for fostering creativity and encouraging innovative thinking. Osborn invented brainstorming for the first time in the advertising company in 1939. Late in 1957, he organized this approach to problem-solving. Applied Imagination approach [80,81]. Following then, brainstorming sparked a global interest in academics and business. People from different ethnic backgrounds get together during the brainstorming phase to work together and communicate in order to produce

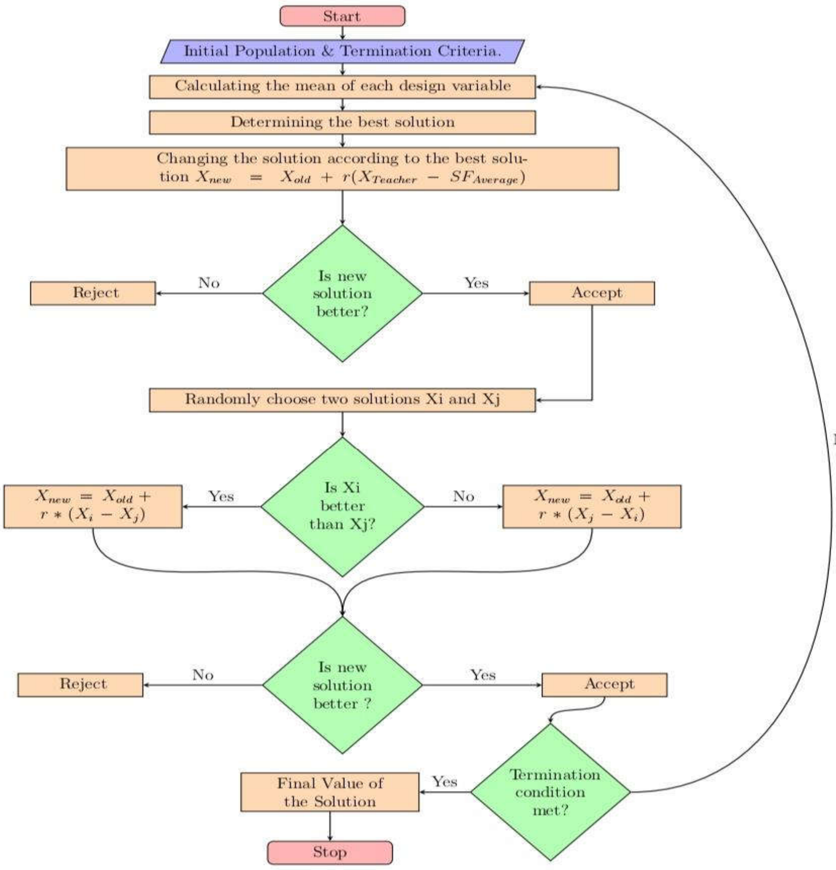


Figure 4.6: Flow Chart Of TLBO Algorithm

fantastic concepts for resolving issues. Algorithm 2 provides the BFOA process stages that were created using brainstorming as inspiration.

Algorithm Six-step Brain Storming Optimization Algorithm explanation:

1: Get going

2: Individuals who might be possible solutions are produced.

3: There are n individuals m groups.

4: N people assessed.

5: Each cluster's members are ranked; the cluster's center is chosen from among the top members.

Sixth: A value 0 and one is produced at random.

Select a cluster center at random if value of a generated is smaller than the value of P5a.

ii Create a fictitious person to represent selected cluster center.

7: Create new persons.

A value is created at random, ranging from 0 to 1.

Select a randomized set a with the probability P6 generated value is smaller than P6b.

ii Produce a haphazard number between 0 and one.

iii) Choose the cluster center and add a random number to create additional people if value is smaller than the pre value of P6biii). iv. If not, select random person from cluster and combine them with the value that was created at random to create new people.

c If not, choose two clusters at random to create new people.
I produce a random number.
ii Choose two cluster centers, combine them, and add randomly produced value create additional people if the produced value is smaller than predefined probability of P6c.
iii If not, two people are chosen at random from each cluster to merge, and the created value is added to create new people.

8: Proceed step 9 n new people are formed; otherwise, proceed step 7. 9: The predefined maximum number repetitions have been achieved; proceed to step three if not. 10: Stop.

4.3.6 Algorithm for Group Leaders' Optimization

An evolutionary algorithm called the Optimization Algorithm (GLOA) was created inspiration from impact of social group leaders. The issue space is split up into various groups, and a leader is chosen for group [82]. Individuals in every group They don't have to similar characters can be made at random. Each group mark its best member to the leader. Every iteration, members of each group want to look like their leaders. The method establishes a solution space in this manner between the group members and the leader. Following a few actions, it was noted this group members top a leader-like appearance. One person is picked at random to promote variety within the group. The variables the other members take the place of some its variables.

Furthermore, a cross over the operator aids in the group's arrival at the local min and a second search of the solution can be conducted to boost variety [83]. Figure 4.7 shows stages that create n groups P individuals and identify group leaders based on their suitability values.

4.3.7 Hierarchically Social Algorithm

social behaviors seen in variety of biological systems and human organizations serve as the basis for the Hierarchical Social Algorithm (HSA). Applications of this meta-heuristic technique to problems with infinite resources, like DFG, have shown success. crucial circuit computation and timing [84]. Simultaneous optimization of the collection of appropriate solutions is the fundamental concept of HSA. Every social group has a workable solution, and these groupings are first dispersed at random to create distinct locations for the solutions. group uses development tactics to compete their neighbors or raise the object function. In this instance, relevant social rivalry and collaboration yield a superior outcome 20. Consequently, the goal solution is maximized. The optimal answer is identified in a single group at the of the procedure [13].

4.3.8 Human Grouping Formation Algorithm

The (HGFA) is a modern social-based meta-heuristic optimization algorithm that actions of both out-group members and in-group members who strive to stay as close to their groups as possible.

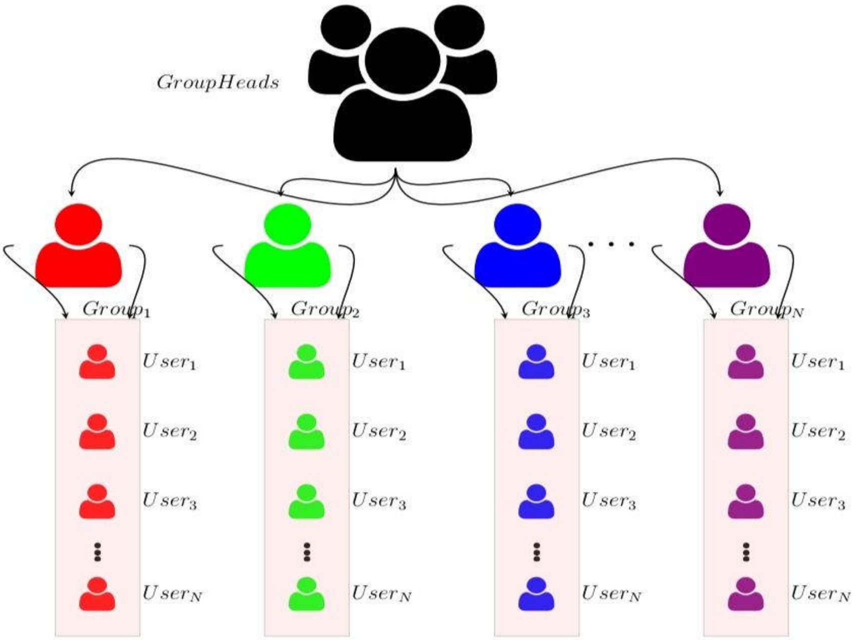


Figure 4.7: Group Leaders Optimization Algorithm

social safety among members of the outgroup [85]. In order characterize human social category, sociologists have the group and out-group condition. Members of in-group are people that the group accepts as

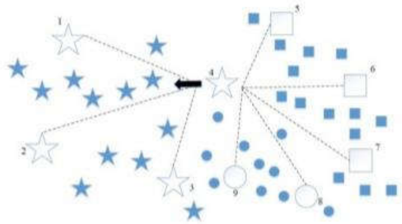


Figure 4.8: Human Group Formation Algorithm

which they are part of. People who identify as belonging to a group acquire that group identity and see themselves as distinct from other groups. They believe their group is better than other groups. Because of this, even while they apart from the group, members the group make every effort to keep group together [86]. It demonstrates how the ideas presented in Figure 4.8 are translated into practical uses.

4.3.9 Social Based Algorithm

An Imperialist Competitor Algorithm-based socio-political process and an evolutionary algorithm are combined to create a new algorithm known as a Social Based Algorithm (SBA). Individuals reside in several kinds of communities: Republic, Autocracy, Monarchy, and international. Additionally, every town has a unique style of leadership. This strategy aims to include a small number of individuals in the characteristic of community development [87]. The phases in the SBA procedure are given in Algorithm 3.

4.4 Social Theory: Outcome Analysis of Benchmark Community Detection Algorithm

Performance evaluation have been done across six distinct graphical media data is sets, including Word adjacencies, rs Zachary Karate [50], Dolphin media, and others, to determine discovery algorithms. Les Misérables, American college football, US politics books, and network [51]

Algorithm 7 A Step-by-Step Guide to Social Based Algorithms

1: Launch

2: Importing the settings

3. Identifying the optimization issue and creating a random population

(c) Choosing a small group of powerful people at random to be leaders; (d) assigning the other people to various regions at random; (e) establishing empires using the imperialist cost function T.Pci;

(f) Choosing charismatic leaders to rule like empires,

4: Ten circles Nd is equal Nd+1.

5: for each and i = 1, 2,. 3.., N, select (a), cross (b), mutation (c), and replacement (d).

6: i = 1, 2,.3.4., N (a) leaders of group are relocated their empire as part the program of human assimilation. (i) x U (0, internal x d) (ii) d: distance between and imperialist

People's Revolution (b) Assimilation Policies of Nations: Each group's leaders settle into their empires, and each nation's citizens follow suit (i) x U (0, coefficient assimilation x) (ii) d: Distance between leader, imperialist.

(d) The national revolution

(f) imperialist racer; selecting the weak nation from the pure empire and awarding it to empire most likely possess it (e) shifting the site

(g) Elimination; the abolition of empire and the powerless principle

7: Verifying the termination requirement and continuing from steps 4 through 7 until it is satisfied.

8: Come to an end.

[52] over normalized information and modularity assessment parameters. Modularity is a structural metric for networks that assesses how well a subgraph (groups, clusters, or communities) inside a network extracts its community structure.

[53]. In a net, clusters nodes with more modularity are comparatively dense with one another, which causes communities to emerge in that network as follows:

$$M = \frac{1}{2|E|}\sum_{xy}[e_{xy} - \frac{w_x w_y}{2|E|}]\delta(c_x, c_y) = \sum_{i=1}^{n}(f_{ii} - f_i'^2) \quad (4.1)$$

Where e_{xy} represents the edge from node x to node y, W_x represent the summation of the weights of the edges linked to node x, c_x is the belonging community structure of node x,$(c_x\ ,\ c_y\ )$ is a probabilistic function that equals to 1 if both the respective node x and y belong to same community structure, otherwise 0. fii represent the edge in community i and F_i' is the belonging probability of random edge to community i that attached to vertices in community i.

Whereas, Normalized mutual information is a normalization of intra-community mutual information score to scale the similarity between intra community node as:

$$nmi(x, c) = \begin{cases} 0 & node\ are\ totally\ dissimilar \\ 1 & node\ are\ totally\ similar \end{cases} \quad (4.2)$$

and mutual information is calculated as

$$nmi(x, c) = \frac{2 * i(x, c_i)}{e(x) + e(c)} \quad (4.3)$$

where the class label is denoted by x, the community structure by c, the entropy bye, and the information gain for element ci for class label x is represented by i(x;c).Benchmark community detection algorithm's performance assessment using and

Tables 4.1 and 4.2 present Modularity and Normalized Mutual information, respectively, without social theories. The integration of social theories with the community detection algorithm results in a considerable improvement of both assessment parameters. SO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO community recognition algorithms yield around 33.26%, 30.76%, 41.95%, 43.05%, 43.14%, 53.13%, 43.02%, and 51.37% modularity, as well as 72.11%, 86.18%, 71.53%, 82.43%, 81.02%, and 86.51%,

Table 4.1: Comparative Analysis of Impact of Social theory on Modularity

Classification Technique	Modularity					
	ZKC	ACF	DCN	BUP	LM	WA
BSO	0.3326	0.6213	0.4844	0.5674	0.5105	0.3611
GLOA	0.3076	0.5947	0.4597	0.5175	0.5098	0.3492
HSA	0.4195	0.5929	0.5981	0.5601	0.5861	0.3030
HGFA	0.4305	0.7203	0.5744	0.6768	0.6079	0.3532
SBA	0.4314	0.6014	0.5655	0.6157	0.6071	0.4187
SEOA	0.5313	0.6107	0.6610	0.6151	0.6121	0.4118
ICA	0.4302	0.7211	0.6179	0.6714	0.6122	0.4192
TLBO	0.5137	0.6204	0.6196	0.7119	0.6217	0.5103

Table 4.2: Comparative Analysis of Impact of Social theory on Normalized Mutual Information

Classification Technique	Normalized Mutual Information					
	ZKC	ACF	DCN	BUP	LM	WA
BSO	0.7212	0.5025	0.5345	0.4225	0.3107	0.3915
GLOA	0.8611	0.6107	0.6227	0.5209	0.3254	0.3927
HSA	0.8118	0.8628	0.7937	0.5261	0.4255	0.3284
HGFA	0.7153	0.6221	0.5755	0.5253	0.4156	0.4582
SBA	0.8243	0.7431	0.7941	0.5143	0.4173	0.4301
SEOA	0.8102	0.8233	0.7162	0.6955	0.5253	0.5822
ICA	0.8651	0.6324	0.5861	0.5712	0.4128	0.4268
TLBO	0.8624	0.7823	0.7152	0.5251	0.4462	0.5122

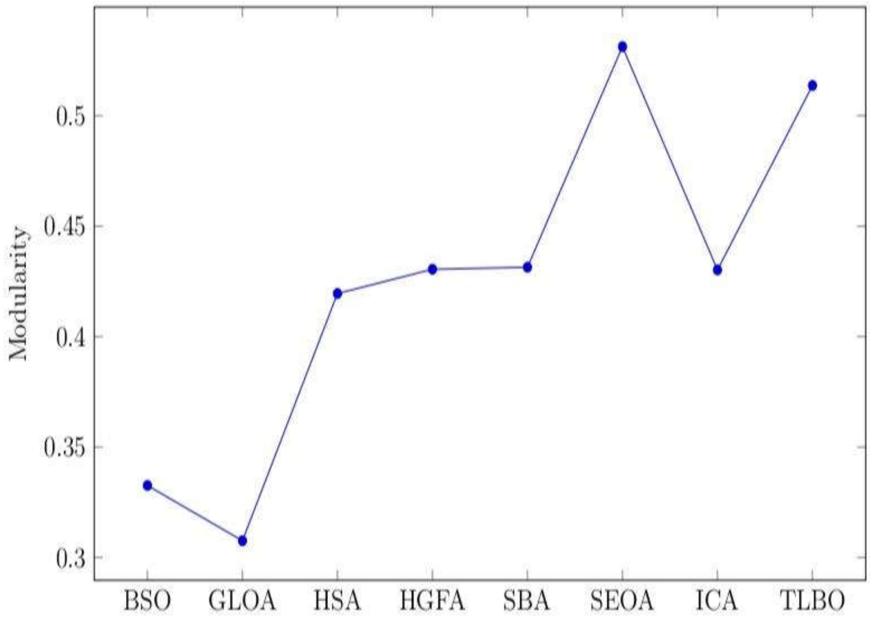


Figure 4.9: Modularity of Community Detection Over ZKC Data Set

Figures 4.9 and 4.10 display, respectively, 86.24% NMI over ZKC data sets. The maximum NMI information is achieved by the TLBO and ICA algorithms, but the SEOA method leads in modularity.

According to figures 4.11 and 4.12, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain approximately 62.13%, 59.47%, 59.29%, 72.03%, 60.14%, 61.07%, 72.11%, and 62.04% modularity over the AFC data set, while the NMI increases by 50.25%, 61.07%, 86.28%, 62.21%, 74.31%, 82.33%, 63.24%, and 78.23%, respectively. The highest NMI information is achieved by the SEOA and HSA algorithms, while the ICA algorithm leads in modularity. On the other hand, as illustrated in figures 4.13 and 4.14, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain approximately 48.44%, 45.97%, 59.81%, 57.44%, 56.55%, 66.10%, 61.79%, and 61.96% modularity and 53.45%, 62.27%, 79.37%, 57.55%, 79.41%, 71.62%, 58.61%, and 71.52% NMI, respectively, over the DCN data set. The highest NMI information is achieved by the SBA and HSA algorithms, but the SEOA algorithm leads in modularity.

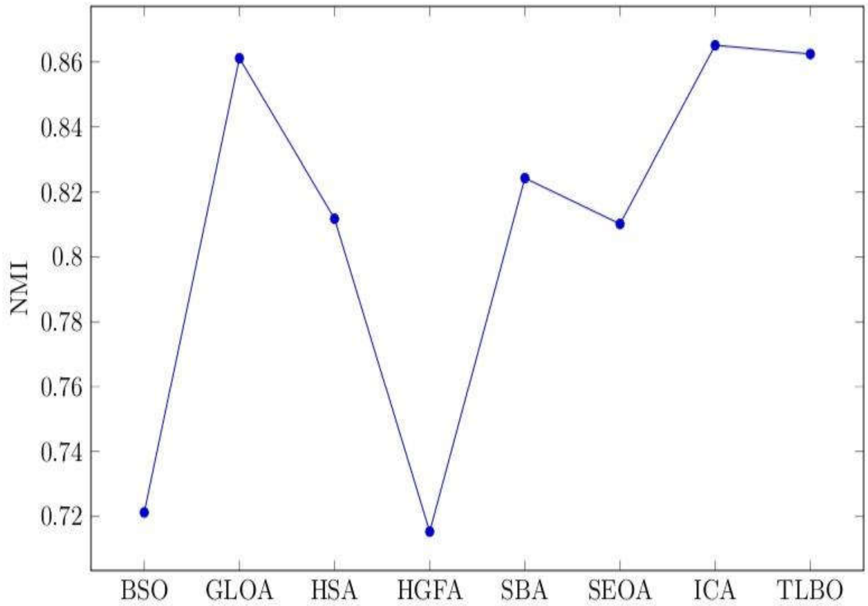


Figure 4.10: Normalized Mutual Information of Community Detection Over ZKC Data Set

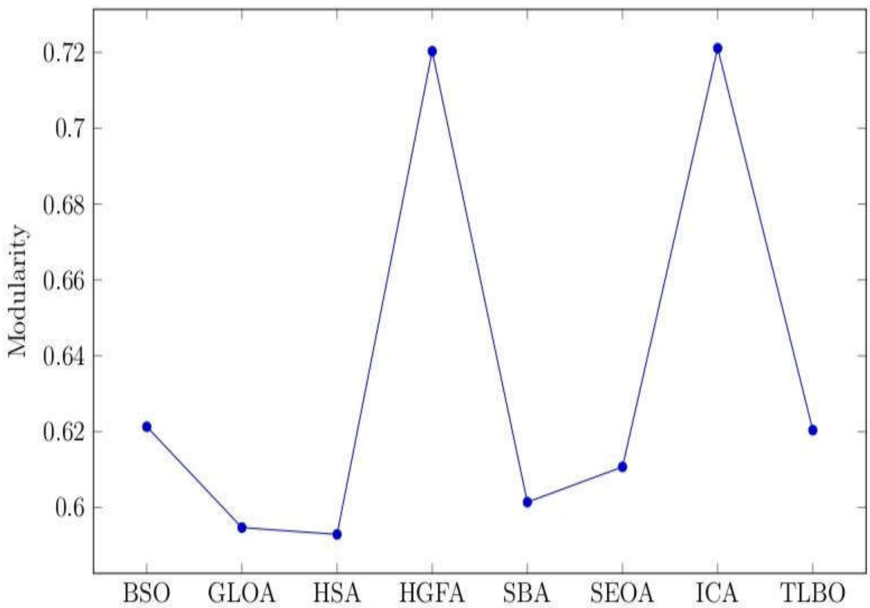


Figure 4.11: Modularity of Community Detection Over AFC Data Set

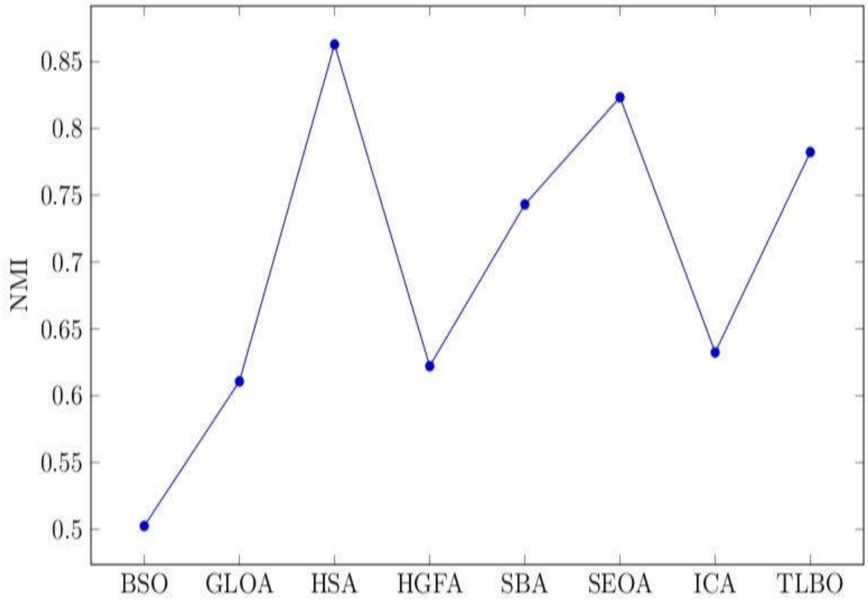


Figure 4.12: Normalized Mutual Information of Community Detection Over AFC Data Set

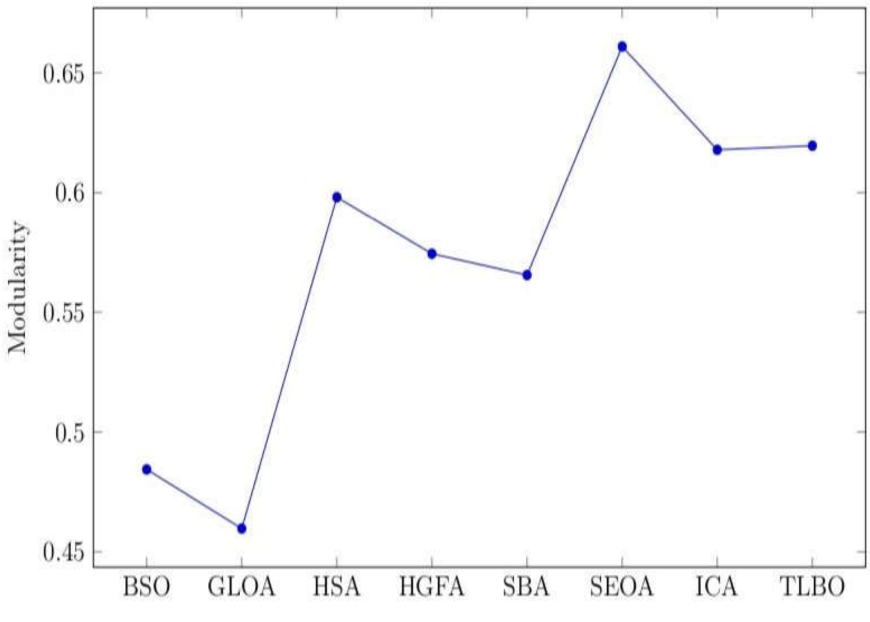


Figure 4.13: Modularity of Community Detection Over DCN Data Set

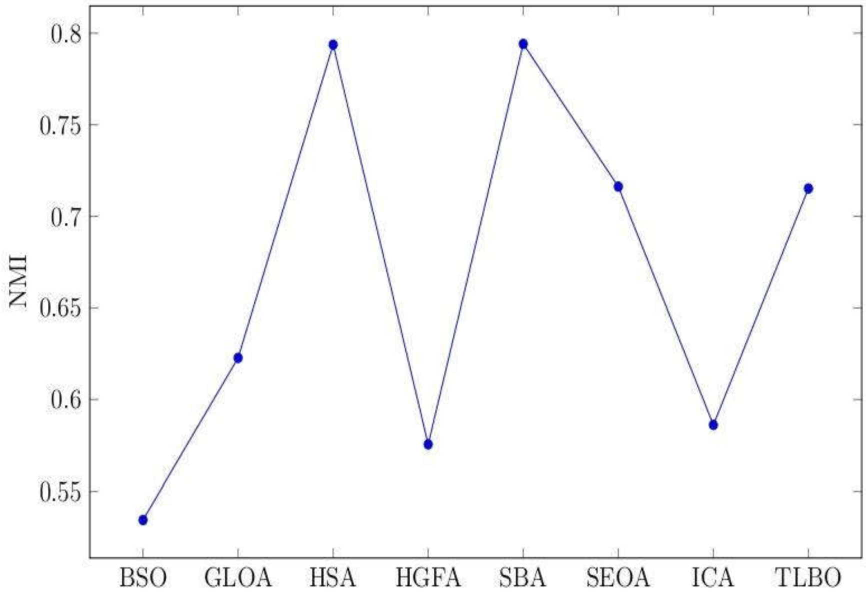


Figure 4.14: Normalized Mutual Information of Community Detection Over DCN Data Set

GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO, and other community discovery algorithms over the BUP data set yield around 56.74%, 51.75%, 56.01%, 67.68%, 61.57%, 61.51%, 67.14%, and 71.19% modularity, while 42.25%, 52.09%, 52.61%, and 52.53%, The NMI is 51.43%, 69.55%, 57.12%, and 52.51%, respectively, as Figures 4.15 and 4.16 illustrate.

The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. Over the LM data set, however, as illustrated in figures 4.17 and 4.18, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain approximately 51.05%, 50.98%, 58.61%, 60.79%, 60.71%, 61.21%, 61.22%, and 62.17% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, and 52.51% NMI,, respectively.

The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. On the other hand, as figure 4.19 and 4.20 illustrate, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain approximately 36.11%, 34.92%, 30.30%, 35.32%, 41.87%, 41.18%, 41.92%, and 51.03% modularity and 39.15%, 39.27%, 32.84%, 45.82%, 43.01%, 58.22%, 42.68%, and 51.22% NMI, respectively, over the WA data set.

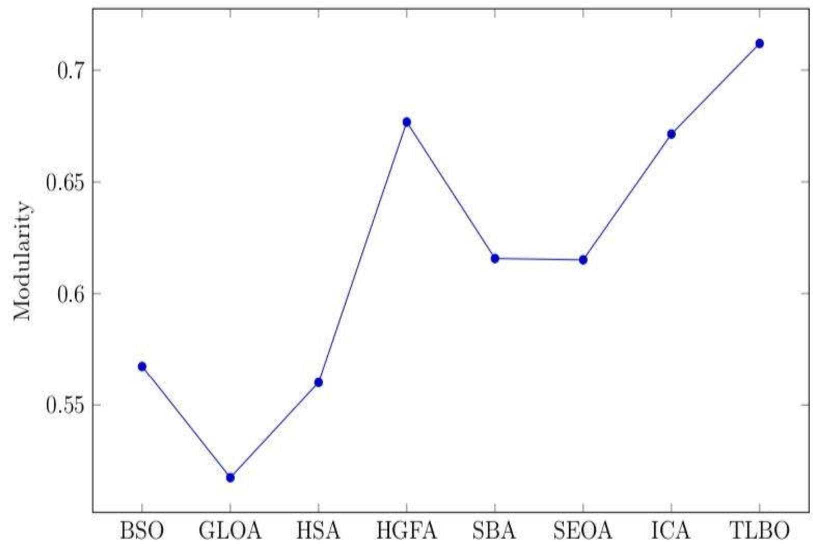


Figure 4.15: Modularity of Community Detection Over BUP Data Set

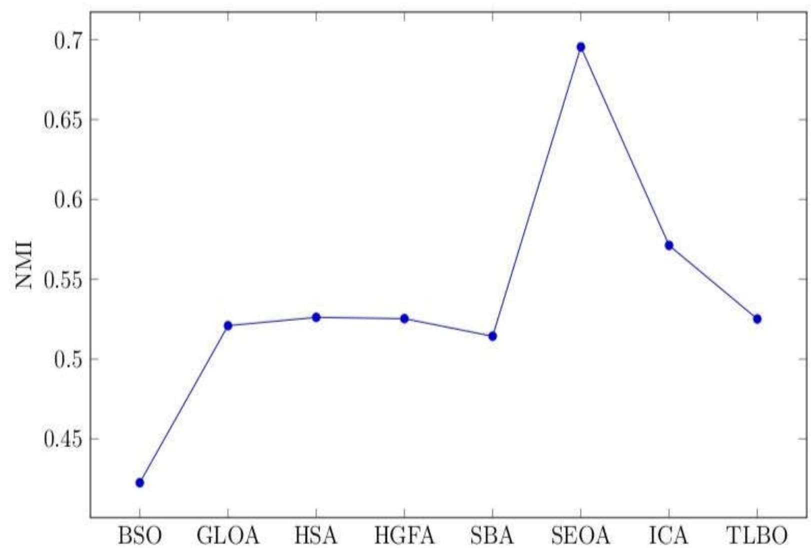


Figure 4.16: Normalized Mutual Information of Community Detection Over BUP Data Set

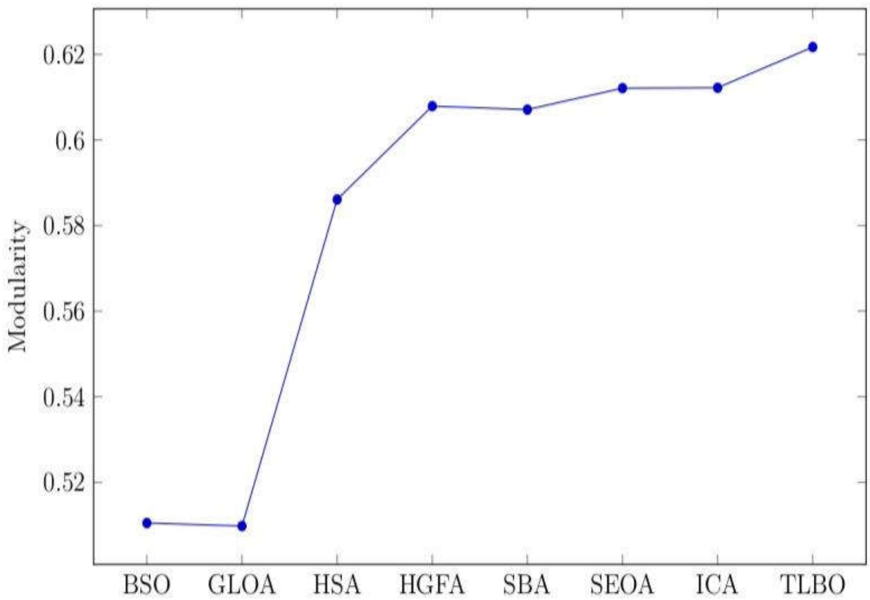


Figure 4.17: Modularity of Community Detection Over LM Data Set

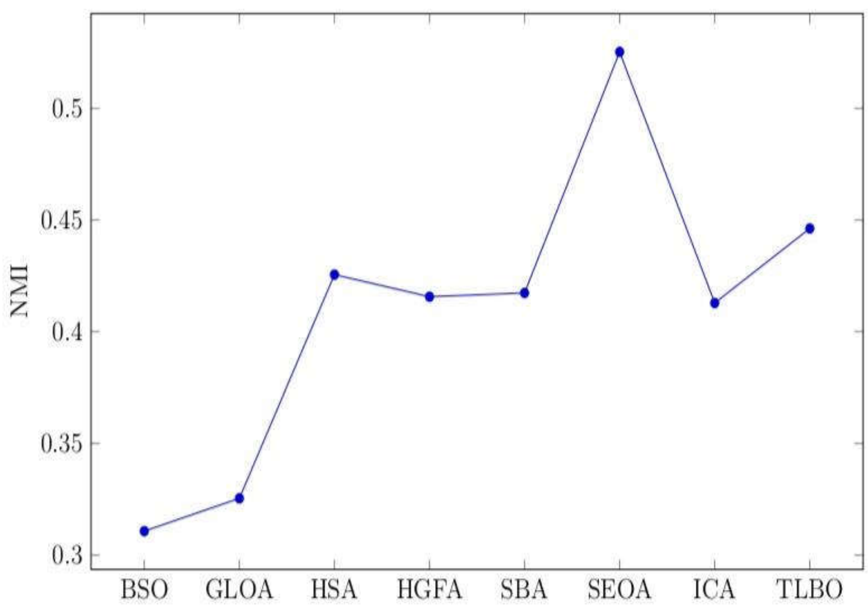


Figure 4.18: Normalized Mutual Information of Community Detection Over LM Data Set

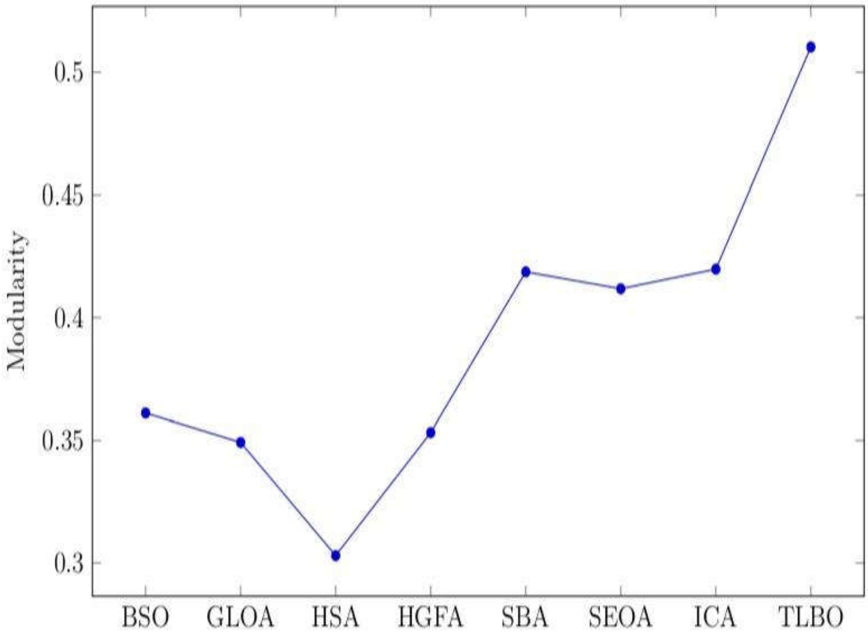


Figure 4.19: Modularity of Community Detection Over WA Data Set

The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. The efficacy of the BSO, GLOA, HSA, and HGFA community detection algorithms Network density affects SBA, SEOA, ICA, and TLBO over social media data sets. It achieves comparatively lower over weakly packed WA data set, greater performance rate, and more dense ACF network.

This part attempts to address community discovery that coincides with a technique that hasn't been used previously and give a thorough assessment on overlapping community structure on social networks, which is commonly experienced in everyday life.

The findings of the investigations and analyses indicate that the methods created to identify overlapping communities in social networks address this issue by focusing on a single goal. This article also provides a comparative examination of six distinct social media-based data sets using meta-heuristic overlapping community recognition techniques. This part examined how the performance rates of the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO varied with the density of the social media data set. It was found that the algorithms attain greater performance rates in dense ACF networks and comparatively lower rates in weakly packed WA data sets.

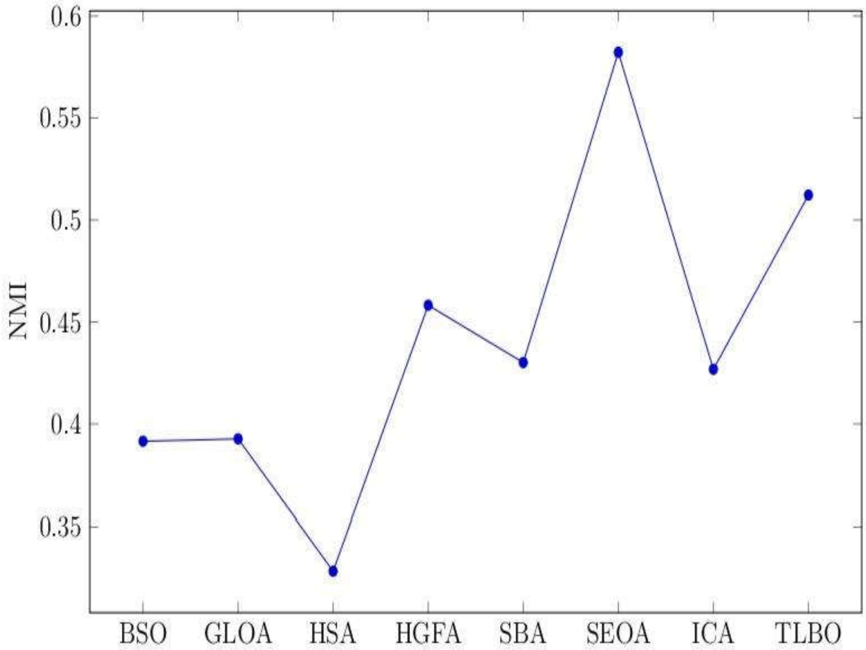


Figure 4.20: Normalized Mutual Information of Community Detection Over WA Data Set

TLBO and ICA are able to extract a more informative community from a denser network than before. In addition, the less dense networks produce superior outcomes for SEOA and HSA.

Chapter 5
Parliamentary Optimization Framework with Multi-Purpose Functions for Social Media Community Detection

5.1 Introduction

The (WWW) exists to give users access to information via websites. Users passive and this presentation one-way on Web 1.0 sites. Some engaged users create and distribute content on Web two.0 sites. Web 2.0 resources that make use Social media refers to platforms for social interaction. Users exchange and discuss their experiences and expertise in these contexts via social sites, blogs, wikis, and discussion boards [13, 88]. Participation, transparency, dialogue, community, and connectedness are some the common characteristics social media [14, 15]. Social networking are online resources that users interact other users inside a certain system and set up profiles. Their profile presentation shows how interact with other users (friends, fans, and followers, for example); it facilitates meeting new people and facilitating conversations with them. Because of the development of virtual communities on these platforms, user performances serve as the foundation for mobility and engagement, which promotes the spread of knowledge. People's efforts to express selves in social networks portray themselves, arouse interest about lives of others, and share motives emerge in all spheres daily life [15–17, 54].Findings from studies on overlapping communities [19, 21, 22] issue in social networking, it have been observed that many the algorithms created earlier for community detection use single purpose to tackle this issue. Simultaneously, a number of recently identified and introduced social-based algorithms have been identified [23, 24, 38, 89]. Parliamentary Optimization (POA) [89] is one these algorithms. Parliamentary elections are simulated by POA [90]. The algorithm's optimization procedure starts with the generation of individual population. These people are regarded as members of parliamentary. The population then divided up among several political parties, and a number of exceptionally qualified individuals are chosen to run for office. Following the orientation , the candidate leading group members are recalculated, with the primary members of group moving in the direction of candidate members. The groups' strengths are determined using the computed new candidates , principal members. In order to prevent their strength from diminishing, strong groups band together, while weaker groupings are eliminated. The population's best individual are solution the optimization issue is accepted when the algorithm's termination condition are satisfied. Within the parameters of this study project, the recently presented POA was employed. This particular process has nevertheless been applied to the issue overlapping community finding social networks before. Using both single and multipurpose functions, community member be included another district using the method that is being employed for the 1 time in this study to find communities that are overlap social networks [38].

The development a new technique allowed social networks to be modular by utilizing a single-purpose component. A fictitious dataset has been used test the proposed methodology. Subsequently, an algorithm designed for one purpose was modified to serve several purposes by incorporating a new goal function to maximize internal density inside network communities. Thus, the first solution for the constructed by combining POA and multifunctional optimization.

5.2 Parliamentary Optimizations Algorithm (POA)

Parliamentarism is another name for the legislative branch of government that creates and enforces laws. In general elections, the public chooses the members of Parliament. Voting for one's favorite party is common. Legislative members who are In parliamentary elections, party members back their respective organizations. In the rivalry between parties, parliamentary groups of members based the party to which they belong aim to surpass other parties. Political parties make up the parliamentary area population in nearly all democracies. Parliamentary elections held under two different systems: proportional representation and majority election. In balanced representation system, many members may chosen from a single constituency, but each constituency. Voters can select which political parties support once each political party has presented its slate of candidates. Parliamentary seats are awarded to parties based on percentage of votes they get [91]. Different political party members, whether they are in or out of parliament, value power differently.

With little influence, these party members try to leave a positive impression other honorable members. This go to this length to win their approval and votes in next elections. Important party members participate in races look for support from the nobles. However, Noble members vote for they trust and are often more resourceful. High capacity general arti are swapped out for prior candidates throughout this procedure. Individuals inside the party compete in this portion of the contest. There is another algorithmic race going on between the parties. Parties want to gain more authority. Obtaining the most members in the legislature and seizing power are the two primary goals of a successful party [91].The first phase in the optimization of Parliamentary Optimization (POA) is to create the starting counting of people. The people that are produced are regarded as parliamentarians. The following stage involves splitting the population up into political parties, after which the candidate with the best fitness for each of the predetermined number member groups is taken into consideration. The in grouping competition begins after this stage. The leading approach the candidate who are a good fit for them during the ingroup competition phase. This scenario is represented by the weighted average primary vectors. candidates [91], as displayed in the POA chart, which is provided in figure 5.1. Several individuals with highest qualification chosen as final candidates is select each group at the conclusion of the group competition process. The next action, the final candidates face off against those from other groups. The group's principal and potential members are crucial in establishing the group's overall influence. Group competition begins after intra-group process competition phase. Parliamentary political parties with one two another to support their nominees.

In order improve their odds of success, strong groups occasionally come together and form one cohesive entity. Algorithm 1 illustrates the POA process phases.

Algorithm 8: Parliamentary Optimization Algorithm (POA) step-by-step explanation:

1: Get going

2: Create the first population.

3. (a) There are M groups made up of L people in the population.

(b) The most physically fit person is chosen as the group's candidate.

(c) Fourth in-group competition:

4. (a) Prominent members approach each group's candidate members.

(b) Appointing new candidates.

(c) Determine each group's power.

5: Group competition

1. The most powerful group is identified, and PM probability is applied to these groupings.

2. Pd, the weakest group, will probably be eliminated.

6: Step 3 repeated if termination condition not satisfied.

7: The optimal choice is thought to be the answer to the optimization issue.

8: Come to an end.

5.2.1 Population Initiation

The d-dimensional issue space is traversed by the N dimension in it solution population at random points. Equation 5.1 shows how d-dimensional continuous are vector is used to code each member of the population.

$$m = [m1, m2,...., mn] \qquad ,mi \ \varepsilon \ IR \qquad (5.1)$$

Every person is the primary member or contender of the designated group. Strengths of individuals are computed based the established fitness function.

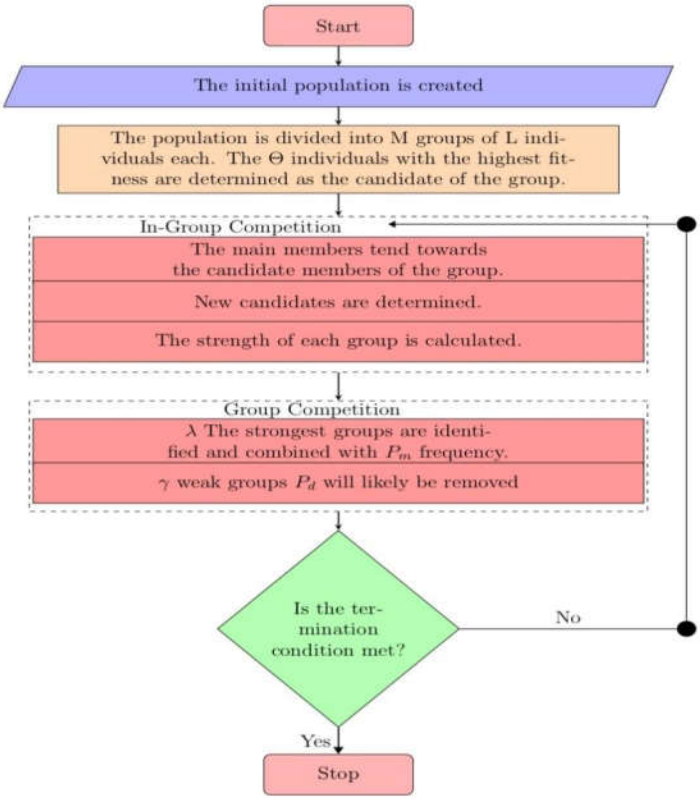


Figure 5.1: Flowchart of Parliamentary Optimization Algorithm

Table 5.1 Values of variables in equation 2

(A.) C1	(B)C2	(B.) C3
(C.) C4	C5	(D.) C6

Table 5.1 Values of parameters in the intra-group competition step

(E.) C1	(B)C2	(F.) C3
(G.) C4	C5	(H.) C6

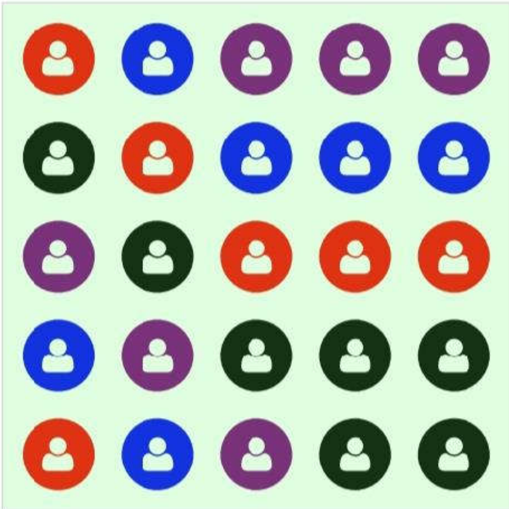


Figure 5.2: Segmentation of the population

5.2.2 Population Segmentation

The population split into as M groups L people to create initial groupings.

$$N = M - L \tag{5.2}$$

N is same as in Equation 5.2, where the positive integers M, L, and N are. superior fidelity $\Theta < L$. Three people are selected to be the groups' candidates. Every group in this algorithmic phase has the same number members. Because of the merger , collapse process, groups may get varying amounts of persons during the algorithm's execution. The split of initial population to three groups five candidates each is depicted in Figure 5.2. The group's candidates are shown figure 5.2 as solid blue symbols.

5.2.3 In-Group Competition

Following the exchange of members between the leading and candidate groups, the primary members of group move in the direction of candidates. The vectors linking a member the orienting process have weighted averages that are precisely proportionate to

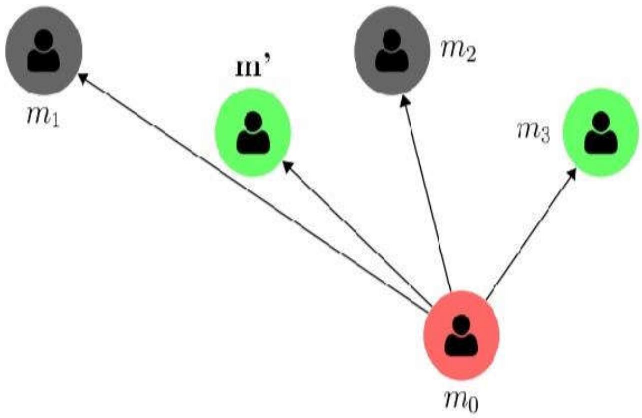


Figure 5.3: Orientation Mechanism

shown in Equation 5.3.

$$m' = m_0 + \eta \left(\frac{\sum_{i=0}^{\theta} (m_i - m_0) . f(m_i)}{\sum_{i=0}^{\theta} f(m_i)} \right) \tag{5.3}$$

finalists. To improve each candidate's candidacy eligibility, each one is weighted accordingly

m′ = m0 + η

η is random number in formula that ranges from 0.51 to 1, enabling algorithm to look for candidates in local are search space. Only in cases when there is high fitness value a primary member replaced. Following the referral procedure, the highest level of fitness members is more than number of potential members. The method for orienting is shown Figure 5.3. While mi is candidate member, m0 is full member. The new leading member position is denoted by m′.The community-based applicants' eligibility value vector evi = {eci,1, evi,2,..., evi,\\} and the candidates' eligibility value ; evc i = {evc i,Θ+1, evc i,\+2,..., evc i,L

i. Equation 5.4 is used to determine the group's strength, which includes the key members.

$$\text{Strength } i = \frac{m.Avg(ev_i)+n.Avg(ev_i^c)}{M+n} \quad ; m>=n$$

5.2.4 Competition between Groups

In order to get stronger, strong groups occasionally come together. To complete the merging, random number is created; if number is smaller than pm, group with highest λ number the more influential and is merged into it. In order to preserve the power and lower the value function, the algorithm eliminates the weaker groupings. The two organizations' merging is depicted in Figure 5.4. Similar to combining, groups with minimal power y are discarded if the resulting random number smaller than PD [92].

5.2.5 Termination of the situation

A group emerges victorious from process, and optimal member of group is deemed the answer to optimization issue. There two instances of dismissal: When the highest number iterations is reached, the algorithm ended. attained or if, after a successful repetition, no discernible increase in fitness value seen [92].

5.3 Detection of Multi-Purpose Overlapping Communities With POA (MPPOA)

Objective function are crucial in optimization issues since they allow POA to be utilized for different purposes. Numerous aim functions, particularly for community finding, have been proposed. If it is agreed that G (V, E) a non-directional chart with m = |E| and n = |V | occurs. The group's set nodes is denoted by S, while set's size is indicated by k. The number sides in set S given by k = |S|.

l = | In {(u, v): u ε S, v ε S}, The number of edges inside the set S's bounds is denoted by cS. cS is equal to | {(u, v): u ε S, v /∈ S} | d(u) represents node d's degree. Below are some objective functions that, when used with the provided definitions, quantify a cluster's quality notion [93].

Conductance: A measure that determines the total number of linkages that leave the cluster.

$$f(S) = cS/2.1 + cS \tag{5.5}$$

† Expansion: Provides the mean quantity of external connections for every cluster node.

$$f(S) = cS/k \tag{5.6}$$

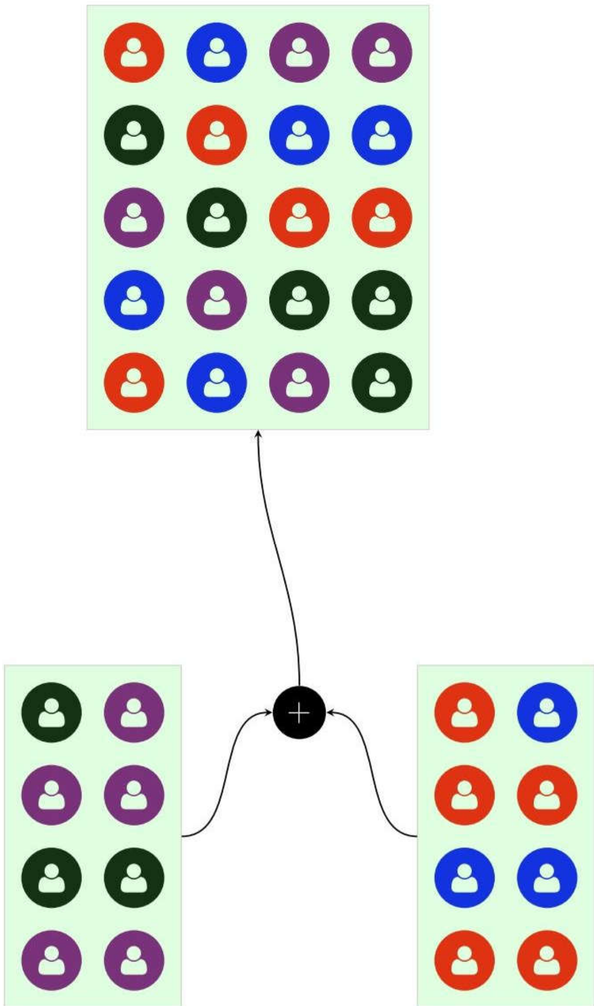


Figure 5.4: Joining Of Group

† Internal Density: The density of the cluster's internal links is represented by S.

$$f(S) = 1 - 1/k\,(k - 1)/2 \quad (5.7)$$

Cut Radio -The percentage of all potential connections departing the cluster is known as .

$$f(S) = 1 - cS/\,k\,(n - k) \quad (5.8)$$

† Standardized Cut: $f(S) = cS/2.1 + cS + cS/\,2.\,(m - 1) + cS$ (5.9)

Maximum − ODF: This represents the proportion of external to internal links for every cluster S node.

Maximum $u \varepsilon S\,|\{(u, v): u /\in S\}|\,/d\,(u)$ 5.10

Average − ODF The average ratio of nodes outside the cluster's connections is represented by † .

$$f(S) = 1\;k\;X\;u\,\varepsilon\,S\,|\{(u, v): u /\in S\}|/\,d\,(u) \quad (5.11)$$

† Flake ODF: This is the proportion of S cluster nodes that has fewer connections outside of cluster than inside of it.

$$f(S) = \backslash\{u: u\,\varepsilon\,S, \backslash\{(u, v): v\,\varepsilon\,S\}\;| < d\,(u)/2\}|\,/k \quad (5.12)$$

As was indicated in the preceding section, the creation initial population marks the beginning of POA. The people in this demographic are regarded as members of parliament.

Those born in the range of 0 to 1 (I1, I2,..., Im) make up the starting population. In Figure 5.5, initial population is expressed.

The entire number nodes the network is represented by matrix as k. The population split in M groups L people each in this stage. At this point, the group's candidates were chosen using a multipurpose technique. Among the goals .

modularity in network serves as a function. A method for identifying overlapping communities has been proposed: expanded modularity [23]. The network's modularity enables the measurement of the caliber of individual segments. compares the distribution links throughout the network to determine the strength the network. Equation 5.13 provides the value extended modularity (ModE).

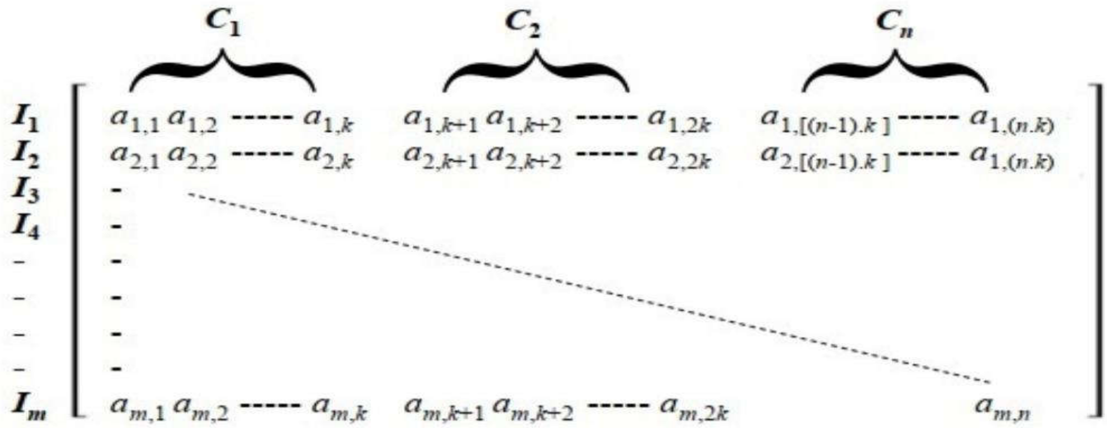


Figure 5.5: Representation of the initial population

$$Mod^E = \frac{1}{2 * \sum e} \sum \ \sum_i \varepsilon cx \ , j \varepsilon cx \ \frac{1}{C_{ni} \ C_{nj}} \left(A_{i,j} \ - \ \frac{d^i \ d^j}{2 * l} \right)$$

The number ensembles to node v belonging is indicated by the equation Cni, and the element the network's neighbor matrix is represented by Ai,j. If there a link bet the i , j nodes, the value of Ai,j is 1; if not, deal 0. di is the node's degree.

P e the network, and i is their sum. Research have shown robust community structures are indicated by a high modularity score. It is therefore anticipated that in this multi-purpose research, the Mod E value would reach its maximum value. This algorithmic step also employs the internal density criteria found in Equation 5.7 as an objective function. One needs to extract a little value from the provided equation.

$$\text{obj}(x) = 1/€n(€e-1)/2 \qquad (5.14)$$

where Pn represents all of the network's nodes. At this POA stage, a multifunctional

Equation 5.15 presents a strategy that combines two objective functions to select the group's candidates.

$$\text{Cost} = a.\text{Mod}^{E} + (1 - a)\,.\text{obj}(x) \qquad (5.15)$$

One the goals is highlighted by using value of an as an input value. As candidates the groups, those with a high cost value the equation admitted. The group's permanent members to candidate members the group competition phase. Equation 3 states that following the orientation , the candidate key group members are once more identified, and Equation 5.4 determines the group's power. Strong groups merge into one group strengthen them after intra-group competition round. The termination requirement is satisfied if there isn't a discernible rise in algorithm steps. A group emerges victorious from the algorithm, and the top individual inside that group deemed the answer overlapping community finding problem.

5.4 Detailed Examination of the Presented Algorithm

This study presents a multipurpose approach to community discovery using POA, and it tests the algorithm's experimental outcomes on various datasets. POA was conducted with a specific goal in mind [92]. In this section of the research, the Programs with a single objective and multipurpose approaches were developed in the Python environment and used to both synthetic and actual network data.

5.4.1 Single-Use Community Discovery on Artificial Dataset with POA

First and foremost, the data representation for the single-purpose research that presented was decided. Figure 5.5 shows the type data representation for the multi-purpose

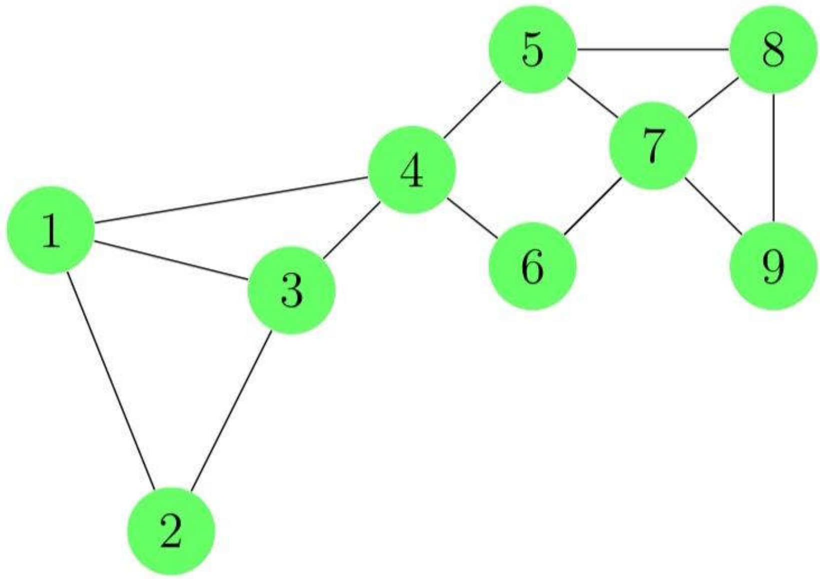


Figure 5.6: A typical network structure

technique is also used to the approach with a single aim. First, the artificial network that was constructed was used to assess the study's effectiveness. The purpose of the project was to gauge POA's capacity for community discovery. A synthetic network made of 13 connections and 9 nodes is shown figure 5.6. First, the initial population in figure 5.7 was generated the algorithm's first stage. Python environment produced these values. Three groups of ten people each comprise the initial population that is created. In this instance, table 5.1 indicates the significance of variables Equation 5.2.The original population generated for the single-objective method is shown Figure 5.7.

5.13: Using an enhanced modularity formula, the distributions the network's connections are computed to determine network's strength. The top three applicants among the determined eligibility values admitted as group members. The

Group competition commences after intragroup competition phase. The strongest $\lambda = 2$ are merged at Pm = 31% or eliminated at Pd ,1% in this stage. Go back to intragroup competition phase if groups decide not to unite. These procedures keep on until all parties are on the same page or the best possible resolution to the issue has been found. Following the combination of the groups, table 5.5 displays the individual eligibility values. The greatest value the table acknowledged as the answer to multipurpose overlapping are community discovery problem, as specified in the POA's termination condition.

Based on table, the first person is regarded as the problem's answer since he is the most fit. Figure 5.7 lists the communities that the POA the synth data set discovered.

	Community1									Community 2								
I_1	0.42	0.15	0.30	0.46	0.57	0.64	0.94	0.46	0.22	0.78	0.81	0.78	0.64	0.80	0.38	0.36	0.33	0.21
I_2	0.36	0.16	0.18	0.56	0.59	0.45	0.95	0.41	0.19	0.73	0.80	0.78	0.57	0.78	0.28	0.46	0.45	0.39
I_3	0.41	0.18	0.26	0.52	0.57	0.57	0.93	0.46	0.22	0.75	0.80	0.79	0.61	0.76	0.36	0.42	0.38	0.28
I_4	0.40	0.18	0.26	0.52	0.57	0.57	0.93	0.46	0.22	0.75	0.81	0.79	0.60	0.75	0.37	0.43	0.37	0.29
I_5	0,42	0,23	0,28	0,55	0.55	0,59	0,91	0,51	0,25	0,70	0,81	0,79	0,59	0,68	0,39	0,47	0,42	0,30
I_6	0,42	0,16	0,25	0,50	0.59	0,57	0,93	0,45	0,21	0,76	0,81	0,79	0,62	0,78	0,35	0,40	0,37	0,26
I_7	0,41	0,17	0,26	0,51	0.58	0,57	0,93	0,46	0,21	0,75	0,80	0,78	0,61	0,76	0,35	0,42	0,38	0,28
I_8	0,43	0,20	0,26	0,52	0.58	0,59	0,91	0,49	0,22	0,73	0,81	0,80	0,60	0,73	0,37	0,44	0,38	0,28
I_9	0,41	0,17	0,26	0,52	0.57	0,57	0,92	0,46	0,22	0,74	0,81	0,79	0,61	0,75	0,36	0,42	0,38	0,28
I_{10}	0,41	0,18	0,25	0,52	0.57	0,58	0,93	0,46	0,21	0,75	0,81	0,79	0,61	0,76	0,35	0,42	0,38	0,28
I_{11}	0,42	0,29	0,42	0,42	0.65	0,56	0,79	0,40	0,24	0,63	0,70	0,66	0,48	0,64	0,40	0,46	0,40	0,41
I_{12}	0,43	0,30	0,41	0,40	0.61	0,57	0,80	0,42	0,26	0,60	0,72	0,67	0,51	0,63	0,42	0,43	0,42	0,37
I_{13}	0,41	0,31	0,43	0,39	0.56	0,57	0,78	0,42	0,32	0,60	0,68	0,72	0,50	0,60	0,42	0,48	0,38	0,37
I_{14}	0,42	0,33	0,39	0,41	0.61	0,56	0,80	0,42	0,27	0,59	0,74	0,67	0,54	0,64	0,43	0,43	0,45	0,37
I_{15}	0,41	0,33	0,41	0,40	0.60	0,56	0,78	0,43	0,28	0,60	0,72	0,67	0,52	0,64	0,44	0,44	0,43	0,40
I_{16}	0,42	0,31	0,41	0,41	0.61	0,56	0,79	0,42	0,27	0,60	0,72	0,68	0,52	0,64	0,42	0,44	0,42	0,39
I_{17}	0,42	0,32	0,40	0,41	0.61	0,57	0,79	0,41	0,26	0,60	0,72	0,67	0,52	0,64	0,42	0,44	0,42	0,39
I_{18}	0,42	0,31	0,41	0,41	0.61	0,56	0,79	0,42	0,27	0,61	0,72	0,67	0,52	0,63	0,43	0,44	0,42	0,39
I_{19}	0,43	0,31	0,40	0,41	0.61	0,57	0,81	0,41	0,28	0,61	0,73	0,68	0,52	0,63	0,42	0,44	0,42	0,38
I_{20}	0,41	0,32	0,41	0,41	0,62	0,56	0,79	0,41	0,26	0,60	0,73	0,67	0,52	0,65	0,43	0,44	0,43	0,40
I_{21}	0,43	0,28	0,37	0,50	0,49	0,55	0,76	0,52	0,30	0,68	0,72	0,76	0,52	0,64	0,31	0,34	0,44	0,39
I_{22}	0,44	0,31	0,37	0,46	0,47	0,57	0,77	0,51	0,34	0,69	0,74	0,76	0,53	0,64	0,34	0,34	0,42	0,35
I_{23}	0,43	0,29	0,37	0,49	0.50	0,54	0,75	0,52	0,31	0,69	0,73	0,76	0,52	0,64	0,32	0,35	0,43	0,39
I_{24}	0,43	0,28	0,37	0,49	0.50	0,55	0,75	0,52	0,31	0,69	0,73	0,75	0,52	0,64	0,32	0,35	0,43	0,34
I_{25}	0,42	0,29	0,36	0,49	0.50	0,55	0,75	0,51	0,31	0,69	0,73	0,74	0,52	0,64	0,31	0,34	0,44	0,39
I_{26}	0,43	0,27	0,37	0,50	0,52	0,53	0,74	0,52	0,29	0,68	0,73	0,76	0,50	0,64	0,32	0,35	0,44	0,42
I_{27}	0,43	0,29	0,37	0,49	0.50	0,54	0,75	0,52	0,31	0,68	0,73	0,76	0,5	0,64	0,32	0,34	0,43	0,39
I_{28}	0,43	0,29	0,37	0,49	0.50	0,55	0,75	0,52	0,31	0,68	0,73	0,76	0,52	0,64	0,32	0,35	0,43	0,39
I_{29}	0,43	0,28	0,37	0,49	0.50	0,54	0,75	0,51	0,31	0,69	0,73	0,75	0,52	0,64	0,32	0,35	0,44	0,39
I_{30}	0,44	0,30	0,37	0,48	0,49	0,55	0,76	0,53	0,32	0,69	0,74	0,76	0,52	0,63	0,33	0,35	0,42	0,37

The values shown in the table that are bolded are acceptable choices for the groupings. Equation 5.3 states that when additional group members approach group candidate, new candidates is chosen. Following the selection of new candidates, each group's strengths were computed, and the results are shown 5.4.

Table 5.3: Eligibility values of Individuals in Groups of the Artificial Data Set.

Group 1		Group 2		Group 3	
I@	EV#	I	EV	I	EV
I_1	1.96	$I_1 1$	1.53	$I_2 1$	1.05
I_2	1.50	$I_1 2$	.80	$I_2 2$	.21
I_3	1.38	$I_1 3$	.86	$I_2 3$	.67
I_4	2.00	$I_1 4$	.65	$I_2 4$	.97
I_5	2.71	$I_1 5$	2.07	$I_2 5$	1.36
I_6	2.90	$I_1 6$	.86	$I_2 6$	2.40
I_7	2.28	$I_1 7$	.26	$I_2 7$	0
I_8	1.50	$I_1 8$	.42	$I_2 8$	1.11
I_9	.61	$I_1 9$	2.07	$I_2 9$	1.25
$I_1 0$	.61	$I_2 0$	2.01	$I_3 0$	1.28

@ Individuals; # Eligibility values

Table 5.4: Strengths of the Groups in the synthetic dataset

Groups	Strengths of the Groups
1.	2.26
2.	1.57
3.	1.41

Table 5.5: Individual Eligibility values in the Artificial Data Set

I@	EV#	I	EV	I	EV
I_1	4.63	$I_1 1$	1.53	$I_2 1$	1.36
I_2	3.78	$I_1 2$	2.07	$I_2 2$	1.28
I_3	3.78	$I_1 3$	.91	$I_2 3$	1.11
I_4	3.78	$I_1 4$	2.07	$I_2 4$	1.36
I_5	2.71	$I_1 5$	3.38	$I_2 5$	1.36
I_6	4.63	$I_1 6$	2.07	$I_2 6$	2.40
I_7	3.78	$I_1 7$	2.07	$I_2 7$	1.11
I_8	3.78	$I_1 8$	1.03	$I_2 8$	1.36
I_9	3.78	$I_1 9$	2.07	$I_2 9$	1.36
$I_1 0$	3.78	$I_2 0$	3.38	$I_3 0$	1.11

@ Individuals; # Eligibility values

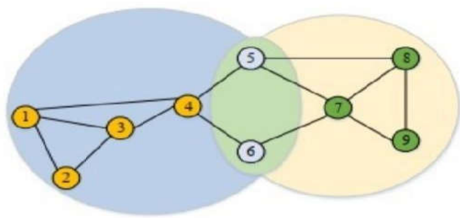


Figure 5.8: Communities found by POA for synthetic dataset

Table 5.6: The communities found for the synthetic dataset.

Communities	Nodes
1	1,2,3,4,5,6
2	5,6,7,8,9

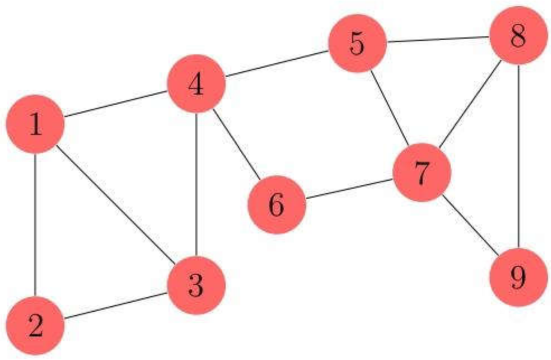


Figure 5.9: Artificial Network created in Pajek Environment

Figure 5.8: POA locates communities for synthetic dataset

Two groups have been discovered by the procedure shown Figure 5.8. Nodes 5 ,6 overlap and are part of two clusters. Table 5.6 lists the communities that the nodes are a part of. Overlapping knots are represented by nodes in bold.

5.4.2 Multifunctional Community Overlap Finding with POA on Synthetic Dataset

An artificial network that was constructed was used to examine the effectiveness of the multipurpose research that was given. The purpose of the project was to gauge POA's capacity for community discovery. A synthetic network with nine nodes and thirteen links is seen in figure 5.9. The algorithm's initial stage was creating the beginning population listed in Table 5.7. The Python environment produced these values. As seen in figure 5.10, the first population generated is split into three groups of ten individuals each. Every member of the group had their fitness values computed using the formula found in Equation 5.15. The top three applicants from the determined eligibility values are admitted into the group. Table 5.7 provides the individual fitness vl for each group. The values shown in the table that are bolded are acceptable choices for the groupings. Equation 3 states that when additional group members approach the group candidate, new candidates are chosen. Following the selection of the new contenders, the

	Community 1									Community 2								
I_1	0,28	0,49	0,34	0,78	0,36	0,73	0,38	0,78	0,67	0,48	0,74	0,64	0,75	0,45	0,56	0,46	0,60	0,34
I_2	0,43	0,74	0,03	0,94	0,76	0,55	0,18	0,49	0,51	0,99	0,85	0,96	0,67	0,40	0,93	0,47	0,23	0,39
I_3	0,40	0,58	0,34	0,78	0,60	0,65	0,63	0,50	0,43	0,44	0,69	0,65	0,78	0,51	0,67	0,44	0,53	0,29
I_4	0,43	0,45	0,25	0,55	0,41	0,69	0,38	0,55	0,47	0,40	0,61	0,76	0,64	0,47	0,59	0,58	0,44	0,55
I_5	0,37	0,52	0,38	0,72	0,44	0,74	0,61	0,60	0,52	0,53	0,79	0,70	0,61	0,57	0,84	0,44	0,48	0,38
I_6	0,20	0,46	0,13	0,61	0,30	0,77	0,53	0,46	0,42	0,41	0,65	0,52	0,49	0,57	0,65	0,52	0,45	0,42
I_7	0,04	0,04	0,09	0,59	0,24	0,84	0,85	0,96	0,48	0,22	0,22	0,53	0,76	0,34	0,46	0,63	0,91	0,16
I_8	0,41	0,59	0,24	0,75	0,42	0,53	0,52	0,66	0,51	0,60	0,82	0,79	0,53	0,41	0,54	0,46	0,34	0,41
I_9	0,46	0,58	0,22	0,53	0,44	0,78	0,36	0,61	0,68	0,36	0,73	0,79	0,58	0,38	0,76	0,56	0,50	0,57
I_{10}	0,19	0,75	0,34	0,41	0,15	0,81	0,62	0,73	0,80	0,06	0,95	0,49	0,75	0,74	0,83	0,15	0,45	0,61
I_{11}	0,32	0,42	0,44	0,42	0,32	0,34	0,63	0,46	0,65	0,40	0,54	0,64	0,28	0,73	0,72	0,49	0,67	0,57
I_{12}	0,37	0,31	0,60	0,44	0,44	0,36	0,56	0,70	0,65	0,49	0,35	0,55	0,20	0,87	0,75	0,66	0,72	0,60
I_{13}	0,40	0,00	0,54	0,20	0,21	0,32	0,09	0,74	0,74	0,54	0,33	0,83	0,55	0,95	0,89	0,35	0,54	0,34
I_{14}	0,39	0,48	0,63	0,40	0,55	0,25	0,52	0,61	0,66	0,30	0,52	0,66	0,26	0,67	0,63	0,50	0,49	0,60
I_{15}	0,05	0,59	0,16	0,83	0,16	0,50	0,99	0,35	0,04	0,21	0,39	0,33	0,22	0,93	0,68	0,96	0,43	0,94
I_{16}	0,19	0,42	0,64	0,44	0,59	0,49	0,65	0,56	0,66	0,29	0,35	0,48	0,37	0,89	0,65	0,52	0,68	0,50
I_{17}	0,51	0,29	0,65	0,58	0,37	0,39	0,51	0,57	0,67	0,21	0,54	0,65	0,30	0,77	0,74	0,61	0,64	0,55
I_{18}	0,38	0,42	0,95	0,57	0,84	0,27	0,62	0,58	0,96	0,08	0,50	0,52	0,09	0,90	0,88	0,43	0,78	0,14
I_{19}	0,39	0,31	0,53	0,63	0,35	0,34	0,54	0,49	0,66	0,50	0,33	0,45	0,41	0,75	0,87	0,70	0,61	0,58
I_{20}	0,32	0,43	0,70	0,54	0,59	0,47	0,54	0,59	0,69	0,24	0,40	0,70	0,32	0,84	0,84	0,70	0,61	0,34
I_{21}	0,31	0,22	0,65	0,06	0,27	0,28	0,88	0,44	0,75	0,60	0,78	0,11	0,97	0,84	0,05	0,46	0,32	0,63
I_{22}	0,36	0,45	0,57	0,32	0,27	0,34	0,70	0,59	0,47	0,52	0,45	0,52	0,67	0,53	0,40	0,45	0,43	0,46
I_{23}	0,59	0,52	0,52	0,24	0,15	0,57	0,70	0,52	0,58	0,43	0,54	0,46	0,72	0,32	0,38	0,63	0,53	0,52
I_{24}	0,51	0,30	0,65	0,13	0,22	0,39	0,84	0,48	0,63	0,58	0,51	0,62	0,43	0,57	0,29	0,65	0,43	0,49
I_{25}	0,51	0,57	0,51	0,16	0,15	0,53	0,79	0,68	0,55	0,68	0,47	0,65	0,44	0,38	0,36	0,72	0,61	0,41
I_{26}	0,15	0,84	0,78	0,27	0,22	0,32	0,82	0,82	0,57	0,57	0,28	0,69	0,79	0,44	0,44	0,46	0,27	0,67
I_{27}	0,57	0,56	0,61	0,21	0,35	0,37	0,58	0,52	0,49	0,56	0,33	0,60	0,65	0,61	0,27	0,64	0,64	0,49
I_{28}	0,37	0,58	0,38	0,23	0,44	0,45	0,64	0,49	0,73	0,42	0,41	0,63	0,57	0,51	0,30	0,65	0,37	0,48
I_{29}	0,29	0,53	0,48	0,40	0,21	0,52	0,84	0,47	0,46	0,38	0,52	0,53	0,45	0,57	0,31	0,45	0,38	0,36
I_{30}	0,78	0,09	0,23	0,24	0,10	0,85	0,69	0,73	0,65	0,51	0,32	0,66	0,11	0,14	0,01	0,96	0,97	0,12

Figure 5.10: Multi-objective Algorithm's Initial Population Creation

Table 5.7: Eligibility values of Individuals in Groups in Artificial Data Set for Multi-purpose Approach

1.Group		2.Group		3.Group	
I@	EV#	I	EV	I	EV
I_1	3.13	$I_1 1$	1.47	$I_2 1$	1.65
I_2	3.45	$I_1 2$	1.10	$I_2 2$	2.89
I_3	1.15	$I_1 3$	1.99	$I_2 3$	.41
I_4	.91	$I_1 4$	.85	$I_2 4$	1.44
I_5	1.95	$I_1 5$	1.90	$I_2 5$	1.83
I_6	1.03	$I_1 6$	1.77	$I_2 6$	1.96
I_7	1.47	$I_1 7$	.62	$I_2 7$	1.15
I_8	3.44	$I_1 8$	1.85	$I_2 8$	1.38
I_9	.64	$I_1 9$	1.10	$I_2 9$	1.29
$I_1 0$	2.24	$I_2 0$	1.10	$I_3 0$	1.70

@ Individuals; # Eligibility values

Table 5.8: Strengths of the Groups in the Synthetic Dataset for Multi-purpose Approach

Groups	Strengths of the Groups
1.	2.26
2.	1.57
3.	1.41

Table 5.9: Eligibility values of individuals in the synthetic data set for the multi-purpose approach

Group-1		Group-2		Group-3	
I@	EV#	I	EV	I	EV
I_1	3.14	$I_1 1$	1.05	$I_2 1$	1.44
I_2	3.45	$I_1 2$	2.09	$I_2 2$	1.14
I_3	3.40	$I_1 3$	1.05	$I_2 3$	1.99
I_4	3.41	$I_1 4$	1.05	$I_2 4$	1.14
I_5	3.42	$I_1 5$	1.83	$I_2 5$	1.90
I_6	3.40	$I_1 6$	1.96	$I_2 6$	1.01
I_7	3.39	$I_1 7$	1.05	$I_2 7$	1.11
I_8	3.37	$I_1 8$	1.03	$I_2 8$	1.36
I_9	1.44	$I_1 9$	1.05	$I_2 9$	1.85
$I_1 0$	1.48	$I_2 0$	1.43	$I_3 0$	1.90

@ Individuals; # Eligibility values

The numbers obtained from the calculation each group's strengths are shown Table 5.8.Group competition commences after the intragroup is competition phase. The two strongest groups (λ = 2) are joined Pm equal 30% or removed for P_d in this stage.= 1%. Go back to intragroup competition phase if groups decide not to unite. These procedures keep on until all parties are on the same page or the best possible resolution to the issue has been found. Following the combination of the groups, table 5.9 displays the individual eligibility values. The greatest value in table is acknowledged as the answer to multipurpose community discovery problem, as specified in the POA's termination condition. The table indicates that because he is the most fit, the second person is thought to be the answer to the issue. Figure 5.11 lists the communities that the POA synthetic data set discovered.

Figure 5.11's algorithm has identified two groups. 5 and 6 overlap and are part of both clusters. Table 5.10 lists the communities that the nodes are a part of. Overlapping knots are represented by nodes in bold.

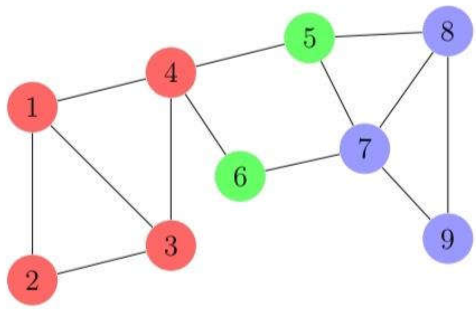


Figure 5.11: Network Structure

Table 5.10: Communities in the Synthetic Dataset for Multi-purpose Approach

Communities	Nodes
1	1,2,3,4,5,6
2	5,6,7,8,9

5.4.3 Discovery of Multipurpose Community Overlap with POA on Real-World Data

This section of the study evaluated the proposed method on four distinct datasets: Lesmis [94], American Football [51], Dolphin Network [52], and Zachary's Karate Club [50]. The goal was to find overlapping groups that might be utilized as a multifunctional strategy.

(I.) The Zachary's Karate Club Data Set is social network of that displays the friendships of 34 members the US University karate club in 1970. The starting population, or first phase of POA, for network with 78 connections was made in the Python environment. There two categories with the first population that is created. The principle and candidate mem of group are identified by computing each member's fitness value using cost value found in Equation 15. The members that are principal candidate are shown Table 5.11. Candidates for membership in the organizations are indicated by names in bold.

(J.) C1	(B)C2	(K.) C3
(L.) C4	(M.) C5	(N.) C6

Following the identification of candidate members, group enters a competitive phase in which permanent members move in the direction of candidate members. Following orientation procedure, the primary members and the candidate the group are computed again. Equation 5.4 is used determine group strengths based the selected applicants and key players. You may get power values in Table 5.12. Group competition commences after intragroup competition phase. The strong $\lambda = 2$ groups are merged Pm = 31% or eliminated at P_d = 1% in this stage. Go back to intragroup competition phase if groups decide not to unite. These procedures keep on until all parties are on the same page or the ideal resolution to the issue has been found. Table 5.13 displays each person's fitness values after all grouper have been pooled. The greatest value in table is acknowledged as the answer to multi- discovery problem, as specified in the POA's termination condition. The ninth person on the table is seen to be the answer to the issue since he is the most fit. Figure 5.12 lists the communities that POA for Zachary's Karate Club discovered. communities are represented by the blue-colored nodes 9, 21, 29, 30, and 31. Table 5.14 lists the communities that the nodes are a part of. In the table, overlapping nodes indicated in bold. (c) Dolphin Network Dataset: 62 dolphins that reside in a confined group in New Zealand make up the nondirectional network of frequent relationships in this dataset, Doubtful Sound. With 159 links in this network, the

In the Python environment, the initial population—the first stage in POA—was produced.

Three groups of ten people each comprise initial population that is created.

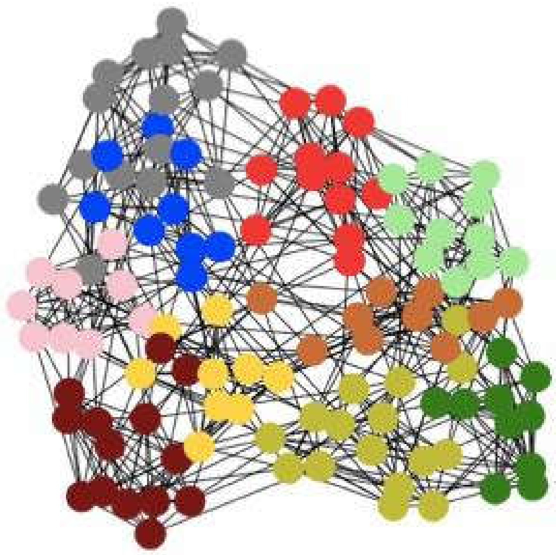


Figure 5.13: American College ball Communities Found POA

(O.) C1	(B)C2	(P.) C3
(Q.) C4	C5	(R.) C6

Table:5.12

(S.) C1	(B)C2	(T.) C3
(U.) C4	(V.) C5	(W.) C6

Table:5.13

(X.) C1	(B)C2	(Y.) C3
(Z.) C4	(AA.) C5	(BB.) C6

Table:5.14

(CC.) C1	(B)C2	(DD.) C3
(EE.) C4	(FF.) C5	(GG.) C6

Table:5.15

Table 5.16 Strengths of the Groups in American College Football Data Set

(HH.) C1	(B)C2	(II.) C3
(JJ.) C4	C5	(KK.) C6

Table 5.17 Eligibility values of individuals in the American College Football data set

(LL.) C1	(B)C2	(MM.) C3
(NN.) C4	C5	(OO.) C6

Table 5.18: Eligibility values of the individuals in the groups in the Dolphin Social Network dataset

Eligibility Values	Individuals									
	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Community-1	129.21	133.80	116.38	113.63	107.21	100.80	94.38	87.96	81.55	75.13
Community-2	65.50	68.26	63.97	64.38	63.62	62.86	62.09	61.33	60.56	59.80
Community-3	102.87	110.92	108.51	113.08	115.90	118.73	121.55	124.37	127.20	130.02
Community-4	143.24	155.66	141.86	145.54	144.85	144.16	143.47	142.78	142.09	141.40
Community-5	65.30	79.61	78.57	87.76	94.40	101.04	107.67	114.31	120.94	127.58
Community-6	103.11	110.92	106.21	109.85	111.40	112.96	114.51	116.06	117.61	119.17
Community-7	91.57	102.72	94.71	99.48	101.04	102.61	104.18	105.75	107.32	108.89
Community-8	111.38	111.95	109.66	109.28	108.43	107.57	106.71	105.86	105.00	104.14
Community-9	133.66	144.16	139.56	145.03	147.98	150.93	153.88	156.83	159.78	162.73
Community-10	190.12	144.85	176.67	157.11	150.38	143.66	136.94	130.22	123.50	116.77

Table 5.19: Strengths of the Groups in Dolphin Social Network Data Set

Groups	Strengths of the Groups
1.	107.40
2.	102.00
3.	130.20
3.	135.20

every one. The principle and candidate members of the group are identified by computing each member's eligibility values in accordance with the cost value found in Equation 5.15. Table presents the principal and candidate members.5.18. The names in bold indicate potential group members.

Following the selection of the candidate members, the group's competitive phase starts, with the permanent members moving in the direction of the candidate members. Following the orientation procedure, a new calculation is made of the applicant and key group members. Equation 5.4 is used to determine the group strengths based on the selected applicants and key players.

Table 5.19 lists the Groups' strengths.

Group competition commences after the intra-group competition phase. The two strongest groups ($\lambda = 2$) are joined at Pm = 30% or eliminated at Pd = 1% in this stage. Go back to the intra-group competition phase if the groups decide not to unite. These procedures keep on until all parties are on the same page or the ideal resolution to the issue has been found. Following the combination of the groups, table 5.20 displays the individual eligibility values. The greatest value in the table is recognized as the answer to the multiple choice question, as specified in the termination condition of the POA.

Table 5.20: Eligibility values of individuals in the Dolphin Social Network dataset

Eligibility Values	Individuals									
	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Community-1	151.18	156.55	136.16	132.95	125.44	117.93	110.43	102.92	95.41	87.90
Community-2	76.64	79.87	74.85	75.33	74.44	73.54	72.65	71.75	70.86	69.96
Community-3	120.35	129.77	126.96	132.30	135.61	138.91	142.21	145.52	148.82	152.12
Community-4	167.60	182.13	165.98	170.29	169.48	168.67	167.86	167.06	166.25	165.44
Community-5	76.40	93.15	91.92	102.68	110.45	118.21	125.97	133.74	141.50	149.27
Community-6	120.64	129.77	124.27	128.53	130.34	132.16	133.98	135.79	137.61	139.43
Community-7	107.14	120.18	110.82	116.39	118.22	120.06	121.90	123.73	125.57	127.41
Community-8	130.31	130.98	128.31	127.86	126.86	125.86	124.86	123.85	122.85	121.85
Community-9	156.39	168.67	163.29	169.69	173.14	176.59	180.04	183.49	186.94	190.39
Community-10	222.44	169.48	206.71	183.81	175.95	168.08	160.22	152.36	144.49	136.63

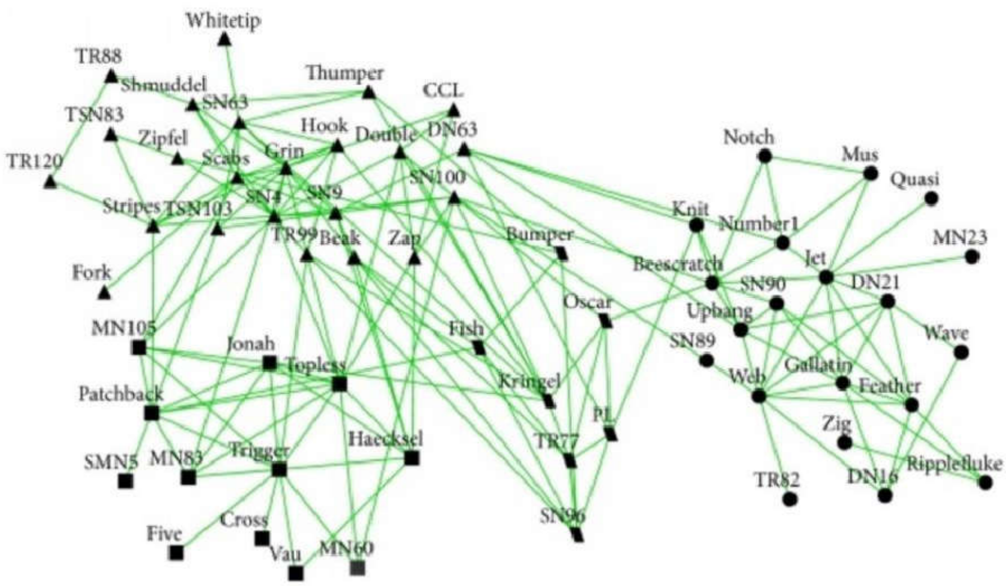


Figure 5.14: Communities found by POA for Dolphin Social Network

difficulty of purpose overlap in community finding.

The table indicates that the second person is seen to be the answer to the issue as he is the most fit. POA discovered communities for Dolphin Social Provided in Figure 5.14 is the network. Two communities has been identified by the Dolphin Social Network algorithm.

These two group1 are represented the green and purple colored nodes, and the overlapping nodes that belong to both communities are shown by the blue colored nodes.

(d) Lesmis Dataset: This dataset depicts the cooperation of 77 characters from Victor Hugo's Les Misérables novel. With 254 links in this network, the Python environment was used build the initial population, which the first stage in the POA process. There are five groups consisting ten people each from initial population. Each group member's eligibility values determined in accordance with

Table 5.21: Eligibility values of the individuals in the groups in the Lesmis data set

Eligibility Values	Individuals in Community									
	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Community-1	232.59	240.84	209.48	204.54	192.99	181.44	169.88	158.33	146.78	135.23
Community-2	117.91	122.88	115.15	115.89	114.52	113.14	111.76	110.39	109.01	107.63
Community-3	185.16	199.65	195.33	203.54	208.62	213.71	218.79	223.87	228.95	234.03
Community-4	257.84	280.20	255.36	261.98	260.74	259.50	258.25	257.01	255.77	254.53
Community-5	117.53	143.31	141.42	157.98	169.92	181.86	193.81	205.75	217.70	229.64
Community-6	185.60	199.65	191.19	197.73	200.53	203.32	206.12	208.91	211.71	214.50
Community-7	164.83	184.89	170.49	179.06	181.88	184.71	187.53	190.36	193.18	196.01
Community-8	200.48	201.51	197.40	196.71	195.17	193.63	192.09	190.54	189.00	187.46
Community-9	240.60	259.50	251.22	261.05	266.36	271.67	276.98	282.29	287.60	292.91
Community-10	342.21	260.74	318.01	282.79	270.69	258.59	246.49	234.39	222.29	210.19

Table 5.22: Strengths of Group in Lesmis data set

Groups	Strengths
1	329.66
2	321.14
3	349.07
4	221.39
5	449.02
6	328.36
7	256.32

The major and potential members of group are identified in relation to the prize value in the Equation 5.15. Table lists the principal candidate members. Candidates for membership in the organizations are indicated by names in bold. Following the selection of candidate members, the group's competitive phase starts, with permanent members moving in the flow of the candidate members. Following the orientation procedure, a new calculation is made of the applicant and key group members. Equation 5.4 is used determine the group strengths based the selected applicants and key players. There are power values Table 5.22. Group competition commences after intragroup competition phase.

The two strongest groups ($\lambda = 2$) are joined at Pm = 31% or eliminated at P_d = 1% in this stage. If groups decide not to combine, the intra-group rivalry will resume.

move. These procedures keep on until all parties are on the same page or the ideal resolution to the issue has been found. When all the categories are joined, table 5.23 provides the individual's eligibility values. As specified in the condition of termination

Table 5.23: Eligibility values of individuals in the Lesmis data set

Eligibility Values	Individuals									
	I-1	I-2	I-3	I-4	I-5	I-6	I-7	I-8	I-9	I-10
Community-1	204.67	211.94	184.35	179.99	169.83	159.66	149.50	139.33	129.17	119.01
Community-2	103.76	108.13	101.34	101.99	100.77	99.56	98.35	97.14	95.93	94.72
Community-3	162.94	175.69	171.89	179.12	183.59	188.06	192.53	197.01	201.48	205.95
Community-4	226.90	246.57	224.71	230.54	229.45	228.36	227.26	226.17	225.08	223.98
Community-5	103.43	126.11	124.45	139.02	149.53	160.04	170.55	181.06	191.57	202.08
Community-6	163.32	175.69	168.24	174.01	176.46	178.92	181.38	183.84	186.30	188.76
Community-7	145.05	162.71	150.03	157.57	160.05	162.54	165.03	167.51	170.00	172.49
Community-8	176.42	177.33	173.71	173.11	171.75	170.39	169.04	167.68	166.32	164.96
Community-9	211.72	228.36	221.07	229.73	234.40	239.07	243.75	248.42	253.09	257.76
Community-10	301.15	229.45	279.85	248.85	238.21	227.56	216.91	206.27	195.62	184.97

The multifunctional overlapping community finding problem's acceptable answer, according to POA, is the value with the greatest value in a table. The ninth person on table is seen to be the answer to the issue since he is the most fit. Figure 5.15 lists the communities that POA for Lesmis was able to locate. Lesmis's recommended algorithm has identified two communities. These two groups are represented the green and purple colored nodes, and the overlap nodes that belong both communities are shown by the blue colored nodes.

5.5 Contextualization and Outcome Evaluation

A thorough computational and experimental environment is designed to analyze the performance the proposed IMCD community detection approach. the Windows-based i7 processor with 4.0 GB of RAM, an integrated graphics card, and a 1TB hard disk Seven experts are employed to carry out these tests. Additionally, the capabilities of RStudio (version 3.2.0) and node4j (version 0.9.1), two open-source software programs, are investigated. Python is used as the programming language for developing the modules.

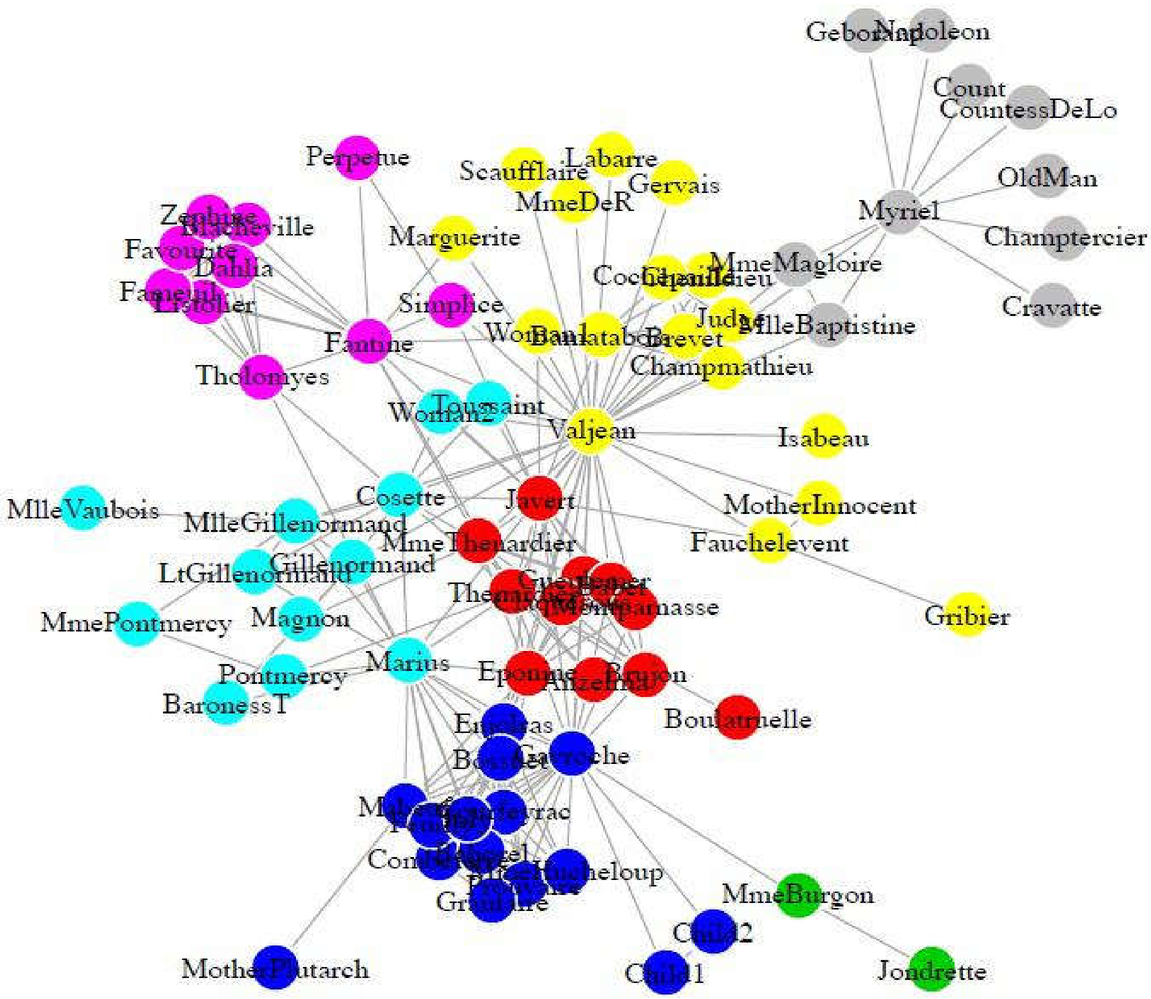


Figure 5.15: Lesmis communities discovered by POA

5.5.1 Tools and Technology

The Windows 7 environment's R studio interface simulates the IMCD framework.

To map actual network data into CSV format, the functionalities of the Neo4j (version 3.3.5) package and the i-graph over R (version 3.3.0) tool are investigated.

5.5.2 Evaluation Measures

influence of a multi-purpose based heuristic community discovery has been assessed through performance assessment on six distinct graphical social media sets, including Word adjacencies, Zachary Karate Club [50], Dolphin social network, and others.

[51], Les Misérables, American politics books, and American college football books [52] over the normalized mutual information and assessment parameter modularity. When assessing the ability of a subgraph (groups, clusters, or communities) inside a network to extract community structure, a structural metric known as modularity is used [53].Communities arise in given network when a set nodes with higher modularity relatively dense with one another. These communities may be described as follows:

Table 5.24: Comparative Analysis of Impact of Social theory on Modularity

Classification Technique	Modularity					
	ZKC	ACF	DCN	BUP	LM	WA
SEOA	0.5313	0.6107	0.6610	0.6151	0.6121	0.4118
ICA	0.4302	0.7211	0.6179	0.6714	0.6122	0.4192
TLBO	0.5137	0.6204	0.6196	0.7119	0.6217	0.5103
MPPOA	0.7456	0.8056	0.7435	0.8123	0.7658	0.6856

Table 5.25: Comparative Analysis of Impact of Social theory on Normalized Mutual Information

Classification Technique	Normalized Mutual Information					
	ZKC	ACF	DCN	BUP	LM	WA
SEOA	0.8102	0.8233	0.7162	0.6955	0.5253	0.5822
ICA	0.8651	0.6324	0.5861	0.5712	0.4128	0.4268
TLBO	0.8624	0.7823	0.7152	0.5251	0.4462	0.5122
MPPOA	0.9087	0.8763	0.8567	0.7652	0.7459	0.7125

and mutual information is calculated as

$$nmi(x, c) = \frac{2 * i(x, c_i)}{e(x) + e(c)} \tag{5.18}$$

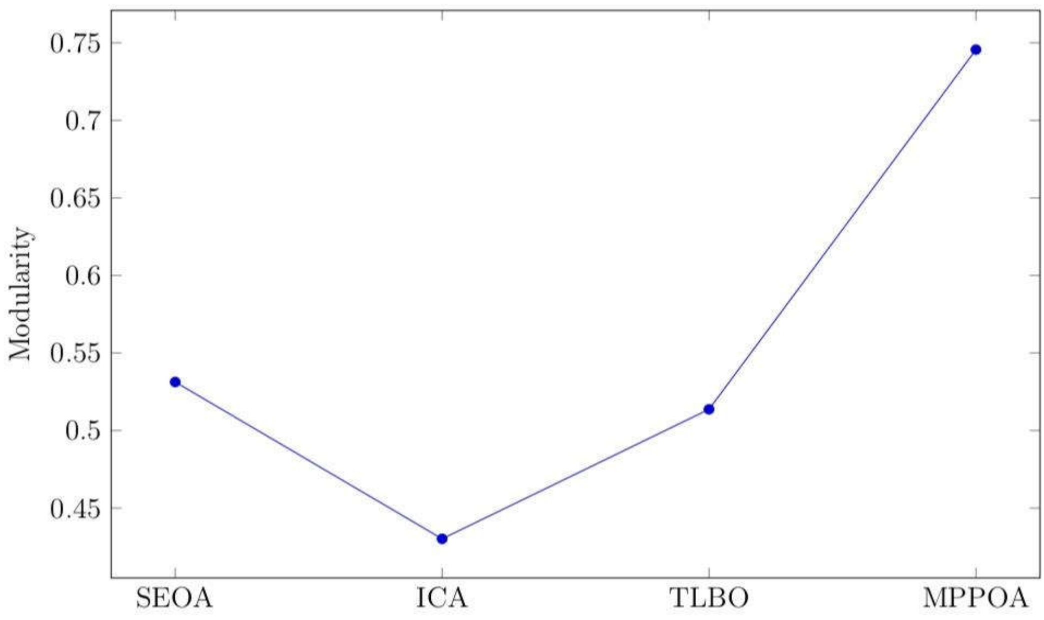


Figure 5.16: Modularity of Community Detection Over ZKC Data Set

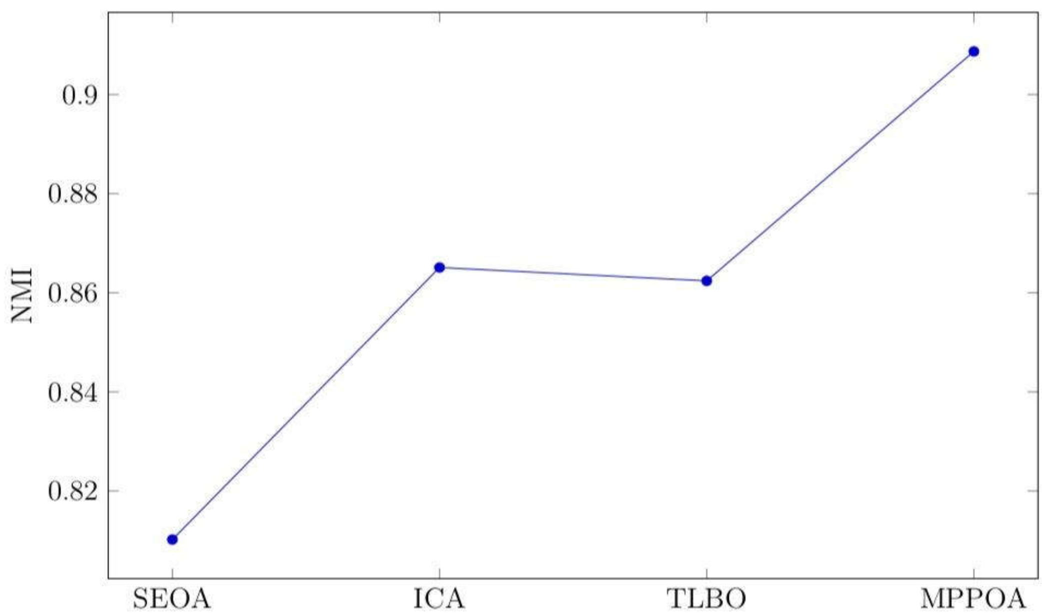


Figure 5.17: Normalized Mutual Information of Community Detection Over ZKC Data Set

Figures 5.18 , 5.19 illustrate how, over the AFC data , community detection algorithms SEOA, ICA, TLB, and MPPOA gain approximately 61.07%, 72.11%, 63.04%, and 81.56% modularity and 83.33%, 63.24%, 78.23%, and 87.63% NMI, respectively. MPPOA

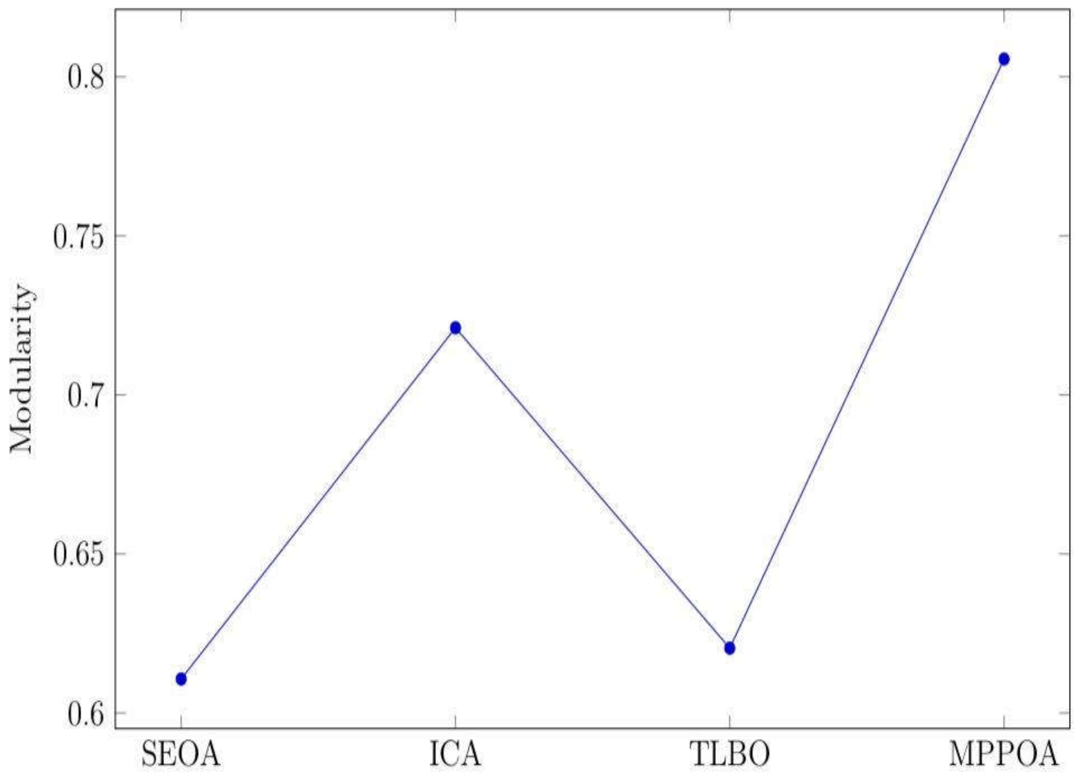


Figure 5.18: Modularity of Community Detection Over AFC Data Set

superior to ICA algorithm in terms of modularity and the NMI information SEOA algorithm, archive was the best prior to MPPOA.

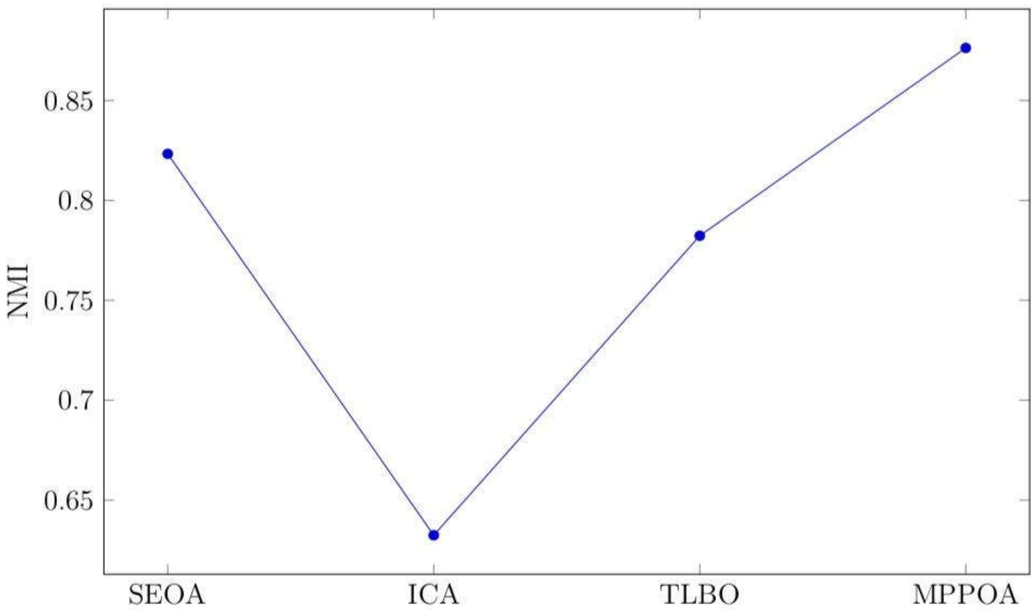


Figure 5.19: Normalized Mutual Information of Community Detection Over AFC Data Set

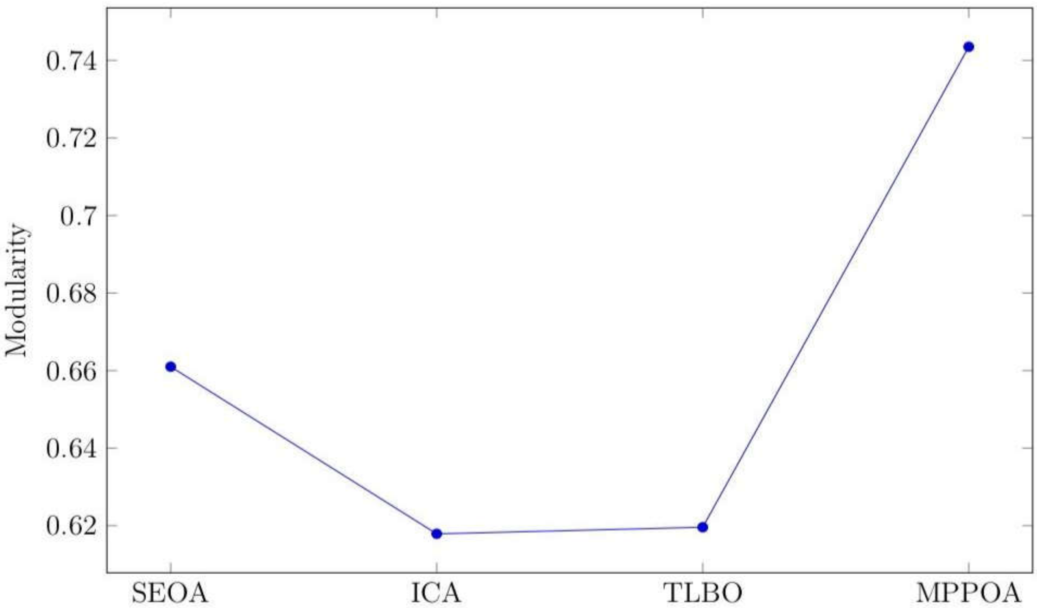


Figure 5.20: Modularity of Community Detection Over DCN Data Set

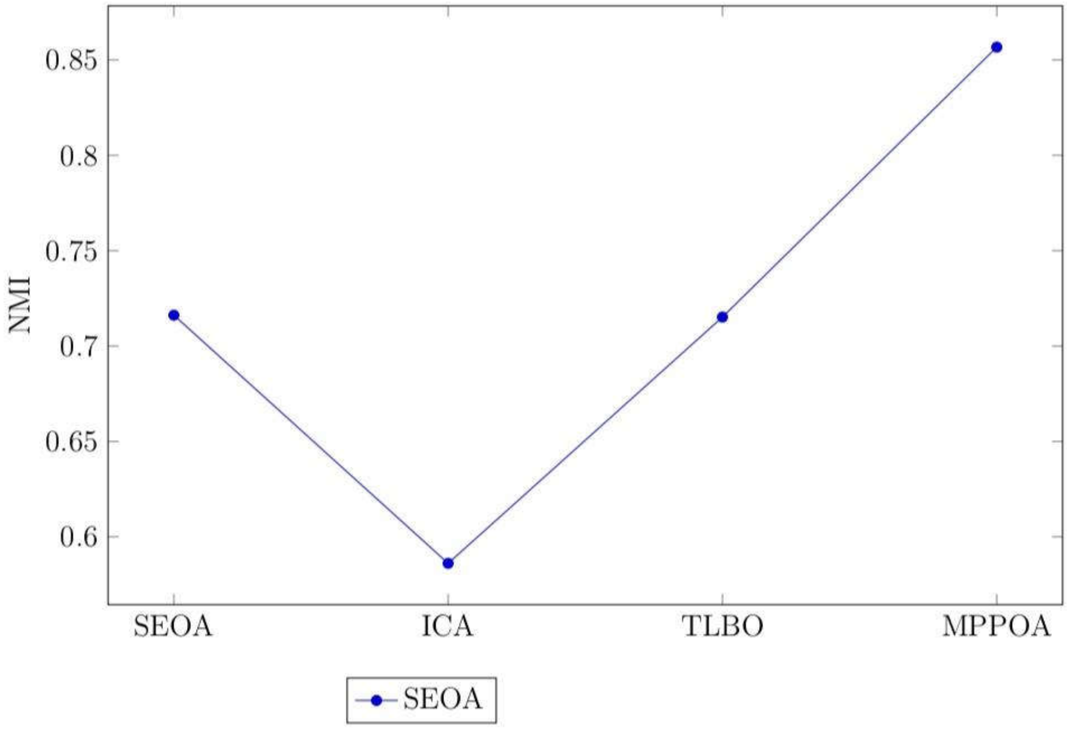


Figure 5.21: Normalized Mutual Information of Community Detection Over DCN Data Set

In contrast, community detection algorithms TLBO ,SEOA, ICA, and MPPOA, as demonstrated in figures 5.20 and 5.21, obtain approximately 66.10%, 61.79%, 62.96%, and 72.35% modularity and 78.41%, 71.62%, 59.61%, 72.52%, and 85.67% NMI, respectively, over the DCN data set. Compared to SEOA algorithm, a whose archive were the best prior to MPPOA, MPPOA performs better.

The high NMI information is achieved by the SBA , HSA algorithms, while the SEOA algo. leads in modularity.

On other hand, as fig 5.22 , 5.23 illustrate, community detection algorithms TLBO,SEOA, ICA, and MPPOA gain approximately 61.51%, 68.14%, 71.19%, 82.23% modularity and 6.855%, 57.12%, 52.51%, and 76.52% NMI over the BUP data set. MPPOA

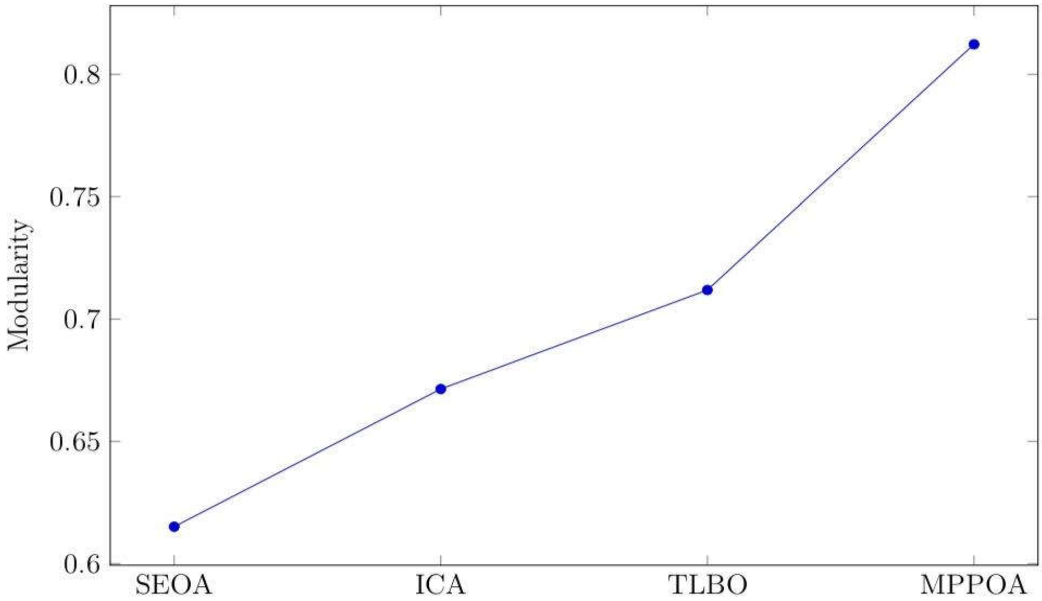


Figure 5.22: Modularity of Community Detection Over BUP Data Set

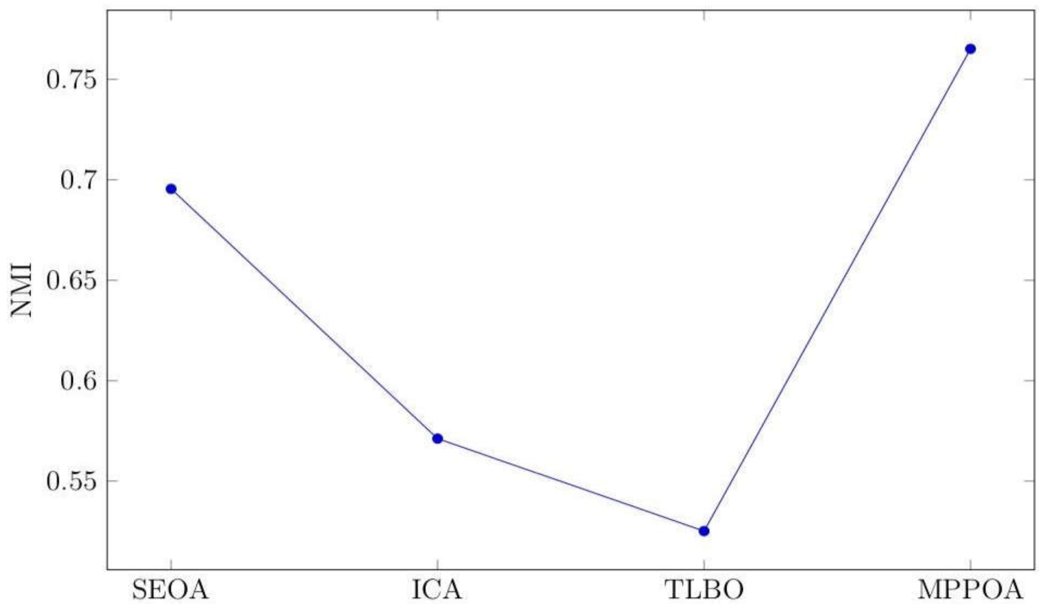


Figure 5.23: Normalized Mutual Information of Community Detection Over BUP Data Set

The modularity of TLBO algorithm is led by the Normalized Information Community Detection DCN Data (Figure 5.21), while SEOA algorithm is led by the NMI information. whose collection was superior to MPPOA's.

Figures 5.24 and 5.25 illustrate how, over the LM data , community detection algorithms ICA, SEOA, TLBO, and MPPOA gain approximately 61.21%, 62.22%, 63.17%, and 75.58% modularity and 67.55%, 57.12%, 53.51%, and 75.59% NMI, respectively. The modularity MPPOA surpasses that of TLBO algorithm, NMI information surpasses that of SEOA algorithm, archive was superior prior to MPPOA.

On other hand, as figure 5.26 and 5.27 illustrate, community detection ICA, SEOA, TLBO, and MPPOA gain approximately 41.18%, 42.92%, 51.03%, 67.56% modularity and 59.22%, 42.67%, 51.22%, and 72.25% NMI over the WA data . TLBO's archive was the best prior to MPPOA, but MPPOA leads in modularity over SEOA , NMI information areover TLBO.

The Overlapping Multi-Purpose Community Detection POA Performance

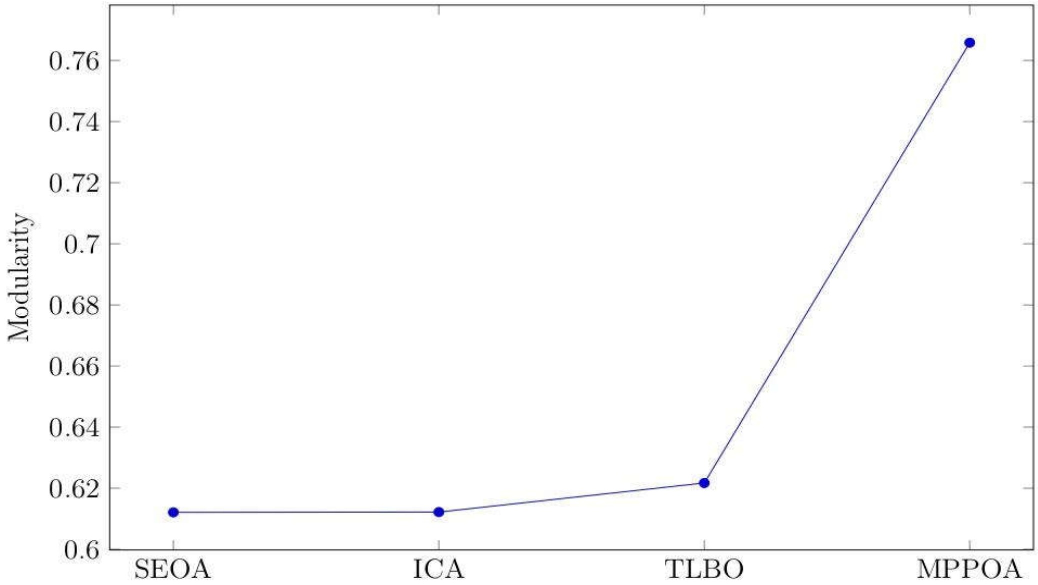


Figure 5.24: Modularity of Community Detection Over LM Data Set

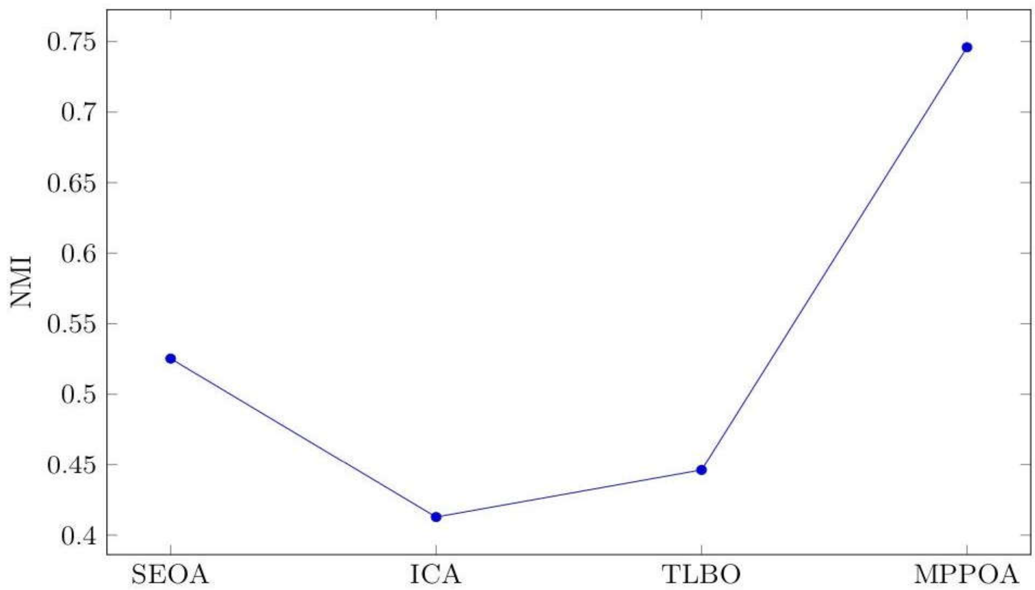


Figure 5.25: Normalized Mutual Information of Community Detection Over LM Data Set

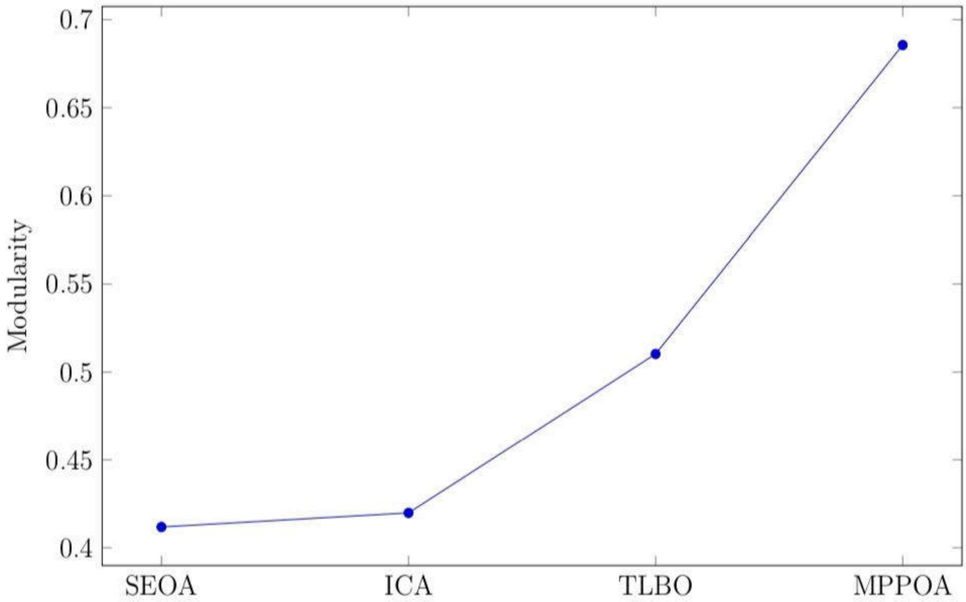


Figure 5.26: Modularity of Community Detection Over WA Data Set

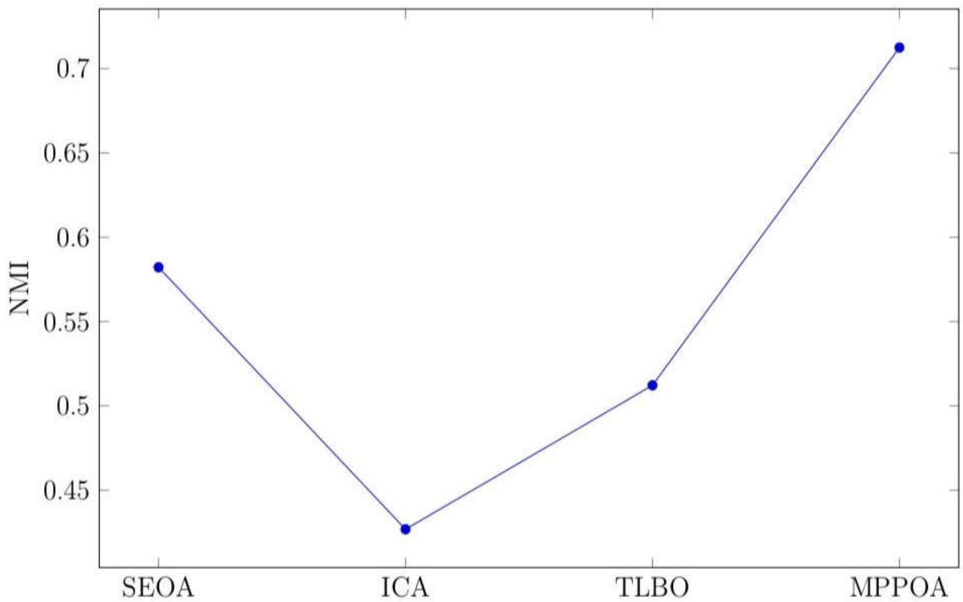


Figure 5.27: Normalized Mutual Information of Community Detection Over WA Data Set

Data set on social media fluctuates according to network density. this attains a comparatively lower overly dense WA data , a higher dense ACF ,ZKC network, and higher performance rate.

Using social media, this chapter introduced Overlapping Multi-Purpose Community Detection are With POA (MPPOA). Significantly, MPPOA organized a very informative and dense community. obtained approximately 92% modularity and 93% NMI across community structure across various data set variations. By comparison, MPPOA outperforms the best-acquired result the combination theory of graphical and social six different real network data sets by 2.24% improvement in modularity and 3.57% improvement in NMI, respectively.

Chapter 6
Conclusions

End users' engagement in corporate marketing, political propaganda, educational activities, entertainment, and commercial activity social media has significantly risen in social media era. End users' inclination foundation has been facilitated by social media.

participation by forming groups of people who have same interests and viewpoints on a range of topics, including local global both politics, products, and concerns. Social media behavior, including likes, dislikes, shares, comments, friends, and profile information, can shape end users' preferences. In order to tackle the delicate character of community structure social media, this thesis focuses on overlapping community finding and the decreased engagement of passive like-minded individuals.

This study solves community discovery that coincides a method has not used previously, and first presents a thorough assessment on overlapping community structure social networks, is commonly experienced in everyday life. It have been shown via investigations and analyses that the methods used to identify overlapping communities in the social networks offer answers to this issue by focusing on single goal. This article also provides a comparative examination of six distinct social media based sets using meta-heuristic overlap community recognition techniques.

community detection algo SEOA,BSO, , HSA, GLOA,HGFA, SBA, ICA, and TLBO were examined in this section in relation to social media sets. The results showed these algorithms achieved greater performance rates in dense ACF networks and comparatively lower rates in lightly dense a WA data sets. Additionally, across the greater dense network, TLBO , ICA are able to extract a higher informative community. The semantic relation-based modularity-optimized community recognition approach for heterogeneous networks is then presented in this paper. This part attempts to use content analysis the network try to raise the modularity ratio of the jointly link analysis and the network. As a result, network was constructed using the similarity values individuals' shares as indirect linkages.

Additionally, this study introduced content- and link-based approaches for the greedy hierarchical clustering algorithm, which makes advantage of indirect connections the network and structure to guarantee that nodes that are closest to one another are prioritized in both topological and semantic groups. The influence semantic relations on optimization algorithms, namely Parliamentary Optimization (POA) and Modularity Optimization (MOA), community discovery is compared in this section. Ultimately, six real network based heterogeneous network information sets were used to test modularity and NMI the study that was given, and the results showed a sufficient modularity rate across the resulting informative community.

Ultimately, this study develops a Parliamentary Optimization (MPPOA) community identification approach based on Multi Purpose functionality that extracts user space for the purpose influencing passive last users. In addition, the semantic relation-based modularity-optimized community recognition technique for heterogeneous networks is presented in this thesis. By combining the network's content analysis with link analysis, this thesis seeks to improve the network's modularity value. As a result, similarity values individuals' shares were computed and added as indirect linkages to the network. Significantly, MPPOA organized a very interesting and dense community. obtained 94% NMI and around 93% modularity throughout the community structure across several data set variations. By comparison, MPPOA outperforms best acquired result benchmark approach with the combination of graphical social theory over 6six distinct actual network data sets by 2.24% improvement in modularity and 3.57% improvement in NMI, respectively.

References

[1] "A guide for choosing community detection algorithms in social network studies: The question alignment approach," N. R. Smith, P. N. Zivich, L. M. Frerichs, J. Moody, and A. E. Aliolo, American Journal of 59th edition, issue 4, pages 597–605, Preventive Medicine, 2020. [Online]. ScienceDirect.com/science/article/pii/S0749379720302166 is the link that can be accessed.

[2] M. Bruglieri and R. Cordone, Computers and Operations Research, vol. 132, p. 105301, 2021, "Metaheuristics for the minimum gap graph partitioning problem." [Online]. Accessible: S0305054821000939 can be found at https://www.sciencedirect.com/science/article/pii

[3] V. Moscato and G. Sperl˩A¬, Knowledge-Based Systems, vol. 224, p. 107112, 2021, "A survey about community detection over on-line social and heterogeneous information networks." [Online]. The following link is accessible: https://www.sciencedirect.com/science/article/pii/S0950705121003750

[4] "Influential nodes detection in dynamic social networks: A survey," N. Hafiene, W. Karoui, and L. Ben Romdhane, Expert Systems with Applications, vol. 159, p. 113642, 2020. [Online]. https://www.sciencedirect.com/science/article/pii/S0957417420304668 is the URL that is accessible.

[5] "Adaptive multi-resolution modularity for detecting communities in networks," S. Chen, Z.-Z. Wang, M.-H. Bao, L. Tang, J. Zhou, J. Xiang, J.-M. Li, and C.-H. Yi, Physica A: Statistical Mechanics and its Applications, vol. 491, pp. 591–603, 2018. Accessible via the following link: https://www.sciencedirect.com/science/article/pii/S0378437117309172

[6] A tabu search-based metaheuristic for dynamic carpooling optimization, S. Ben Cheikh-Graiet, M. Dotoli, and S. Hammadi, Computers and Industrial Engineering, vol. 140, p. 106217, 2020. [Online]. Accessible: S0360835219306862 can be found at https://www.sciencedirect.com/science/article/pii

[7] "A graphbased clustering algorithm for software systems modularization," B. Pourasghar, H. Izadkhah, A. Isazadeh, and S. Lotfi, Information and Software Technology, vol. 133, p. 106469, 2021. Accessible via the following link: https://www.sciencedirect.com/science/article/pii/S0950584920302147

[8] "An improved multi-objective evolutionary algorithm for detecting communities in complex networks with graphlet measure," D. A. Abduljabbar, S. Z. M. Hashim, and R. Sallehuddin, Computer Networks, vol. 169, p. 107070, 2020. [Online]. [https://www.sciencedirect.com/science/article/pii/S138912861931182X] is accessible.

[9] An information entropy and latent dirichlet allocation approach to noise patent filtering is presented by J. Yoon, B. Jeong, M. Kim, and C. Lee in Advanced Engineering Informatics, vol. 47, p. 101243, 2021. [Online]. [https://www.sciencedirect.com/science/article/pii/S1474034620302123] is accessible.

[10] G. Costa and R. Ortale, "Bayesian probabilistic generative modeling and mean-field variational inference: Integrating overlapping community discovery and role analysis," Engineering Applications of Artificial Intelligence, vol. 89, p. 103437, 2020. [Online]. [https://www.sciencedirect.com/science/article/pii/S0952197619303343] is accessible.

[11] Q. Zhou and C. Zhang, Journal of Informetrics, vol. 15, no. 3, p. 101162, 2021, "Breaking community boundary: Comparing academic and social communication preferences regarding global pandemics." [Online]. https://www.sciencedirect.com/science/article/pii/S175115772100033X is the URL that is accessible.

[12] "Podcd: Probabilistic overlapping dynamic community detection," S. Bahadori, H. Zare, and P. Moradi, Expert Systems with Applications, vol. 174, p. 114650, 2021. [Online]. ScientificDirect.com/science/article/pii/S0957417421000919 is the link that can be accessed.

[13] H. Deylami, A. and M. Asadpour, “Social network link prediction using hierarchical neighborhood detection," in the 7th Conference on Information and Knowledge (2015) 2015, Technology (IKT), pp. 1–5.

[14] F. Lotfi, S. Karimi, and H. Izadkhah, "Group-oriented "Link prediction within multiplex networks," Informetrics Journal, v. 15, no. 4, 2021, p. 101178. [Online]. Accessible:/science/article/pii/S1751157721000493 at https://www.sciencedirect.com/

[15] A. Fang, "The impact of opinion dynamics and communication structure" in multi-trusted state social networks," Applied Mathematics and Processing, vol. 2021; 406, p. 126262. [Online]. Accessible:

ScienceDirect.com/science/article/pii/S0096300321003519 has been accessed.

[16] Question. Chen along with L. Wei, "Combining communities to identify intricate networks: Ansurvey," in the 20th International Parallel and Distributed Computing Conference, 2019 Technologies and Applications (PDCAT), 2019, pp. 513–516.

[17] D. N. C. Debnath, P., K. Singh, R. A. Haraty, and N. Choudhury, "Examining the

2020 IEEE, "dynamic community detection algorithms in complex networks."

Pages 989–994 in International Conference on Industrial Technology (ICIT), 2020.

[18] S. Z and Liu. Li, "An adjusted genetic algorithm for detecting communities in intricate

networks," in the International Conference on Methodologies and Algorithms, 2017

2017, pp. 1-3 in Models and Applications in Emerging Technologies (ICAMMAET).

[19] B. Shen, H. and N. Wang. "A novel genetic algorithm for overlapping community," Qiu

identification," in the Tenth International Conference on Intelligent Information held in 20142014, pp. 766–769. Hiding and Multimedia Signal Processing.

[20] M. Guerrero, A. Alcayde, R. Ba^{h}os, F. G. Montoya, and C. Gil, "Using genetic algorithms for adaptive community detection in complex networks," Neurocomputing, section V. 266, 2017, pp. 101–113. [Online]. Accessible:

This link points to a science article with a publication index of S0925231217308664.

[21] K. Fatania, T. Patalia, D. Joshi, and Y. Jejani, "A comparison of communities that overlap"identification in a vast, intricate network, at the 2019 International Conference on ICRAECC, 2019: Recent Developments in Energy-efficient Computing and Communication, pp. 1–6.

[22] "Similarity preserving overlapping community detection in signed networks," C. He, H. Liu, Y. Tang, S. Liu, X. Fei, Q. Cheng, and H. Li, Future Generation Computer Systems, vol. 116, pp. 275–290, 2021. [Online]. Accessible: S0167739X20330181 can be found at https://www.sciencedirect.com/science/article/pii

[23] Detection of hierarchical overlapping communities in complex networks by Y. Yan and H. Xie, in Proceedings of the 2019 International Conference on Intelligent Computing, Automation, and Systems (ICICAS), pp. 893–896.

[24] Performance comparison of physics-based meta-heuristic optimization algorithms by D. DEMIROL, F. OZTEMIZ, and A. KARCI, in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1–5.

[25] "Median evidential c-means algorithm and its application to community detection," K. Zhou, A. Martin, Q. Pan, and Z. ga Liu, Knowledge-Based Systems, vol. 74, pp. 69–88, 2015. [Online]. [http://www.sciencedirect.com/science/article/pii/S095070511400402X] is accessible.

[26] In Social Networks, vol. 40, pp. 154–162, 2015, A. G. Nikolaev, R. Razib, and A. Kucheriya, "On efficient use of entropy centrality for social network analysis and community detection." Accessible via the following link: http://www.sciencedirect.com/science/article/pii/S0378873314000550

[27] "A semantic overlapping community detection algorithm based on field sampling," X. Yu, J. Yang, and Z.-Q. Xie, Expert Systems with Applications, vol. 42, no. 1, pp. 366–375, 2015. [Online]. ScienceDirect.com/science/article/pii/S0957417414004072 is the link that can be accessed.

[28] The article "Similarity between community structures of different online social networks and its impact on underlying community detection" was published in the 2015

issue of Communications in Nonlinear Science and Numerical Simulation by W. Fan and K. Yeung. [Online]. ScienceDirect.com/science/article/pii/S1007570414003165 is the link that can be accessed.

[29] "A parameter-free community detection method based on centrality and dispersion of nodes in complex networks," by Y. Li, C. Jia, and J. Yu Statistical mechanics and its applications, Physica A, vol. 438, 2015, pp. 321–334. [Online]. Accessible: Science Direct: http://www.sciencedirect.com/science/article/pii/S0378437115006032

[30] "Fuzzy duocentric community detection model in social networks," Social Networks, vol. 43, pp. 177–189, 2015, Golsefid, M. H. F. Zarandi, and S. Bastani. [Online]. ScienceDirect.com/science/article/pii/S037887331500043X is the link that can be accessed.

[31] Expert Systems with Applications, vol. 42, no. 21, pp. 8349–8360, 2015. Y. Xu, H. Xu, and D. Zhang, "A novel disjoint community detection algorithm for social networks based on backbone degree and expansion." Accessible via the following link: http://www.sciencedirect.com/science/article/pii/S0957417415004443.

[32] Chen, F. and Li, K., "Detecting hierarchical structure of community members in social networks," Computational Intelligence Applications for Data Science, vol. 87, no. 87, pp. 3–15, 2015. [Online]. [http://www.sciencedirect.com/science/article/pii/S0950705115002129] is accessible.

[33] "Memetic algorithm with simulated annealing strategy and tightness greedy optimization for community detection in networks," Applied Soft Computing, vol. 34, pp. 485–501, 2015, C.-H. Mu, J. Xie, Y. Liu, F. Chen, Y. Liu, and L.-C. Jiao. Accessible via the following link: http://www.sciencedirect.com/science/article/pii/S1568494615003415

[34] Physica A: Statistical Mechanics and its Applications, vol. 436, pp. 430–442, 2015. X. Zhou, Y. Liu, B. Li, and G. Sun, "Multiobjective biogeography based optimization algorithm with decomposition for community detection in dynamic networks." [Online].

ScienceDirect.com/science/article/pii/S0378437115004847 is the link that can be accessed.

[35] Expert Systems with Applications, vol. 84, pp. 74–91, 2017, L. Duan, Y. Liu, W. N. Street, and H. Lu, "Utilizing advances in correlation analysis for community structure detection." [Online]. ScienceDirect.com/science/article/pii/S0957417417303238 is the link that can be accessed.

[36] A. Biswas and B. Biswas, "Regression line dominance analysis of evolutionary optimization and community detection algorithms," Information Sciences, vol. 396, pp. 185–201, 2017. [Online]. Accessible: ScientificDirect.com: http://www.sciencedirect.com/science/article/pii/S0020025517305492

[37] Iedc: An integrated approach for overlapping and non-overlapping community detection, M. Hajiabadi, H. Zare, and H. Bobarshad, Knowledge-Based Systems, vol. 123, pp. 188–199, 2017. [Online]. ScienceDirect.com/science/article/pii/S0950705117300850 is the link that can be accessed.

[38] "A density based link clustering algorithm for overlapping community detection in networks," X. Zhou, Y. Liu, J. Wang, and C. Li, Physica A: Statistical Mechanics and its Applications, vol. 486, pp. 65–78, 2017. [Web]. [http://www.sciencedirect.com/science/article/pii/S0378437117305502] is accessible.

[39] "Deep community detection in topologically incomplete networks," X. Xin, C. Wang, X. Ying, and B. Wang, Physica A: Statistical Mechanics and its Applications, vol. 469, pp. 342–352, 2017. Accessible via the following link: http://www.sciencedirect.com/science/article/pii/S0378437116308342

[40] P. G. Sun, "Imbalance problem in community detection," 16 vol. 457, pp. 364–376, Physica A: Statistical Mechanics and its Applications, 2016. Accessible via the following link: http://www.sciencedirect.com/science/article/pii/S0378437116300838

[41] "Hwmscb: A community-based hybrid approach for identifying influential nodes in the social network," by A. Ahmad, T. Ahmad, and A. Bhatt Statistical Mechanics and its

Applications, Physica A, 2019. [Online]. ScienceDirect.com/science/article/pii/S0378437119319983 is the link that can be accessed. In Future Generation Computer Systems, vol. 94, pp. 173–184, 2019, A. Singh, S. Garg, S. Batra, and N. Kumar, "Probabilistic data structure-based community detection and storage scheme in online social networks," [Online]. ScienceDirect.com/science/article/pii/S0167739X17310300 is the link that can be accessed.

[43] V. Moscato, A. Picariello, and G. Sperla, "Game theory-based community detection," Artificial Intelligence in Engineering Applications 2019; Intelligence, vol. 85, pp. 773–782. [Online]. ScienceDirect.com/science/article/pii/S0952197619301885 is the link that can be accessed.

[44] Clustering and evolutionary approach for longitudinal web traffic analysis: A study by A. Morichetta and M. Mellia, Performance Evaluation, vol. 135, p. 102033, 2019. [Online]. Accessible: ScientificDirect.com/science/article/pii/S0166531619300331

A novel collaborative filtering recommendation approach based on soft co-clustering is presented by M. Li, L. Wen, and F. Chen in Physica A: Statistical Mechanics and its Applications, vol. 561, p. 125140, 2021. Accessible via the following link: https://www.sciencedirect.com/science/article/pii/S0378437120305963

[46] In Physica A: Statistical Mechanics and its Applications, vol. 483, pp. 156–170, 2017, B. Yang, H. He, and X. Hu, "Detecting community structure in networks via consensus dynamics and spatial transformation." Accessible via the following link: https://www.sciencedirect.com/science/article/pii/S0378437117304077.

[47] "Overlapping community detection in rating-based social networks through analyzing topics, ratings, and links," A. Reihanian, M.-R. Feizi-Derakhshi, and H. S. Aghdasi, Pattern Recognition, vol. 81, pp. 370–387, 2018. [48] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: Tfˆa¿"idf, lda, and doc2vec," Information Sciences, vol. 477, pp. 15–29, 2019. Available: https://www.sciencedirect.com/science/article/pii/S0031320318301481. Accessible via

the following link:
https://www.sciencedirect.com/science/article/pii/S0020025518308028.

[49] Layout-based computation of web page similarity ranks by A. S. Bozkir and E. Akcapinar Sezer, International Journal of Human-Computer Studies, vol. 110, pp. 95–114, 2018. [Online]. ScienceDirect.com/science/article/pii/S1071581917301465 is the link that can be accessed.

[50] The article "An information flow model for conflict and fission in small groups" was published in the Journal of Anthropological Research in 1977. The article was written by W. W. Zachary.[Online]. The URL https://doi.org/10.1086/jar.33.4.3629752 is accessible.

[51] D. Lusseau, Proceedings of the Royal Society of London, "The emergent properties of a dolphin social network." S186–S188 in Series B: Biological Sciences, vol. 270, no. suppl 2, 2003.

[52] M. Girvan and M. E. J. Newman, Proceeding of the National Academy of Sciences, vol. 99, pp. 7821–7826, 2002, "Community structure in social and biological networks." The paper "Modularity and community structure in networks" by M. E. J. Newman was published in the Proceedings of the National Academy of Sciences in 2006. The article can be found on pages 8577–8582. [Online]. The following is available: 103/23/8577.

[54] A generalized incremental bottom-up community detection framework for highly dynamic graphs by W. Liu, T. Suzumura, L. Chen, and G. Hu, in 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 3342–3351.

In Online Social Networks and Media, vol. 3–4, pp. 32–48, 2017, G. Rossetti, D. Pedreschi, and F. Giannotti, "Node-centric community discovery: From static to dynamic social network analysis." [Online]. ScientificDirect.com/science/article/pii/S2468696417301052 is the link that can be accessed.

[56] In the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 2497–2500, I. Litou and V. Kalogeraki present "Pythia: A system for online topic discovery of social media posts."

[57] "Social set analysis: A set theoretical approach to big data analytics," R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, IEEE Access, vol. 4, pp. 2542–2571, 2016.

[58] Sphynx: A parallel multi-gpu graph partitioner for distributed memory systems, S. Acer, E. G. Boman, C. A. Glusa, and S. Rajamanickam, Parallel Computing, p. 102769, 2021. [Online]. ScientificDirect.com/science/article/pii/S0167819121000272 is the link that can be accessed.

[59] "Multivariate times series classification through an interpretable representation," Informa- tion, F. J. Bald˛A¡n and J. M. Ben˛Atez Science, volume 569, pages 596–614, 2021. [Online]. https://www.sciencedirect.com/science/article/pii/S0020025521004825 is the URL that is accessible.

[60] Scalable spectral clustering for overlapping community detection in large-scale networks: H. Van Lierde, T. W. S. Chow, and G. Chen, IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 4, pp. 754–767, 2020.

[61] In the 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 1892–1896, H. Jain and G. Harit present "Unsupervised temporal segmentation of human action using community detection."

[62] In 2019 Pattern Recognition Letters, vol. 122, pp. 14–22, published a paper by N. F. Haq, M. Moradi, and Z. J. Wang titled "Community structure detection from networks with weighted modularity." [Online]. https://www.sciencedirect.com/science/article/pii/S0167865519300303 is the URL that is accessible.

[63] A novel approach for overlapping community detection using multi-objective optimization, M. Ebrahimi, M. R. Shahmoradi, Z. Heshmati, and M. Salehi, Physica A: Statistical Mechanics and its Applications, vol. 505, pp. 825–835, 2018. Accessible via

the following link:
https://www.sciencedirect.com/science/article/pii/S0378437118303534.

[64] In 2020, IEEE Access published a paper titled "Overlapping community detection in weighted temporal text networks," written by R. Dong, J. Yang, and Y. Chen.

[65] "Constrained spectral clustering for dynamic community detection," A. Karaaslanli and S. Aviyente, in ICASSP 2020 - 2020 (ICASSP), 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 8474–8478.

[66] Percolation on interdependent networks with cliques and weak interdependence, W. Zang, X. Ji, S. Liu, and G. Wang, Physica A: Statistical Mechanics and its Applications, vol. 566, p. 125612, 2021. [Online]. https://www.sciencedirect.com/science/article/pii/S0378437120309109 is the URL that is accessible.

[67] "Community detection in networks: A multidisciplinary review," by M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig The Journal of Network & Computer Applications, vol. 108, 2018, pages 87–111. [Online]. https://www.sciencedirect.com/science/article/pii/S1084804518300560 is the URL that is accessible.

[68] In 2017 International Conference on Computing Intelligence and Information System (CIIS), Y. Zhou and F. Xu, "Research on application of artificial intelligence algorithm in directed graph," pp. 116–120.

[69] Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis, M. Ehteram, A. Ferdowsi, M. Faramarzpour, A. M. S. Al-Janabi, N. Al-Ansari, N. D. Bokde, and Z. M. Yaseen, Alexandria Engineering Journal, vol. 60, no. 2, pp. 2193–2208, 2021. Accessible via the following link: https://www.sciencedirect.com/science/article/pii/S1110016820306840

[70] Physical Communication, vol. 44, p. 101240, 2021; T. Li, J. Yang, and D. Cui, "Artificial-intelligence-based algorithms in multiaccess edge computing for the

performance optimization control of a benchmark microgrid." [Sciencedirect.com/science/article/pii/S1874490720303177]. Available

[71] In Computers and Structures, vol. 244, p. 106395, 2021, E. Ficarella, L. Lamberti, and S. Degertekin compare three novel hybrid metaheuristic algorithms for structural optimization problems. [Online]. [https://www.sciencedirect.com/science/article/pii/S004579492030198X] is accessible.

[72] An efficient imperialist competitive algorithm with likelihood assimilation for topology, shape, and sizing optimization of truss structures is described by M. Dehghani, M. Mashayekhi, and M. Sharifi in Applied Mathematical Modelling, vol. 93, pp. 1–27, 2021 [Online]. ScientificDirect.com/science/article/pii/S0307904X2030696X is the URL that can be accessed.

[73] A modified teaching-based optimization algorithm for solving optimization problems was developed by Y. Ma, X. Zhang, J. Song, and L. Chen. The paper was published in Knowledge-Based Systems, vol. 212, p. 106599, 2021. [Online]. At https://www.sciencedirect.com/science/article/pii/S0950705120307280, it is accessible [74]. Group teaching optimization algorithm: A novel metaheuristic approach for resolving global optimization problems, Y. Zhang and Z. Jin, Expert 148, p. 113246, Systems with Applications, 2020. [Online]. ScientificDirect.com/science/article/pii/S0957417420300725 is the link that can be accessed.

[75] A. Ammar and S. ul Islam Ahmad, "A research reactor's core reload pattern optimization using a teaching-learning based optimization algorithm," Annals of Nuclear Energy, vol. 133, pp. 169–177, 2019. [Online]. Accessible: This link points to a science article with a publication index of S0306454919302786.

[76] Applied Mathematical Modelling, vol. 77, pp. 309–326, 2020; A. K. Shukla, P. Singh, and M. Vardhan, "An adaptive inertia weight teaching-learning-based optimization algorithm and its applications." [Online]. ScientificDirect.com/science/article/pii/S0307904X19304470 is the link that can be accessed.

[77] "Socio evolution learning optimization algorithm: A socio-inspired optimization methodology," by M. Kumar, A. J. Kulkarni, and S. C. Satapathy, was published in Future Generation Computer Systems, vol. 81, pp. 252-272, 2018. [Online]. ScientificDirect.com/science/article/pii/S0167739X17317259 is the link that can be accessed.

[78] Pattern Recognition Letters, vol. 54, pp. 36–42, 2015; Zhang, Y. and P. Zhang, "Machine training and parameter settings with social emotional optimization algorithm for support vector machine." [Online]. ScientificDirect.com/science/article/pii/S0167865514003614 is the link that can be accessed.

[79] "Social emotional optimization algorithm based identification of nonlinear Hammerstein model," P. S. Pal, S. Choudhury, A. Ghosh, S. Kumar, R. Kar, D. Mandal, and S. P. Ghoshal, in 2016 International Conference on Communication and Signal Processing (ICCSP), 2016, pp. 1633–1637.

[80] "Modified brain storm optimization algorithm in objective space for pollution-routing problem," A. Rauniyar, R. Nath, and P. K. Muhuri, 2019 IEEE Congress on Evolutionary Computation (CEC), 2019, pp. 242-247.

[81] An optima-identified framework with brain storm optimization for multimodal optimization problems is presented by Z. Dai, W. Fang, K. Tang, and Q. Li in Swarm and Evolutionary Computation, vol. 62, p. 100827, 2021. [Online]. ScienceDirect.com/science/article/pii/S2210650220304806 is the link that can be accessed.

[82] L. Jain, R. Katarya, S. Sachdeva, "Whale optimization algorithm in online social network for opinion leader detection," Expert Systems with Applications, vol. 142, p. 113016, 2020. [Online]. Accessible: S095741741930733X, https://www.sciencedirect.com/science/article/pii

[83] "A novel intelligence algorithm based on the social group optimization behaviors," X. Feng, Y. Wang, H. Yu, and F. Luo, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 48, no. 1, pp. 65–76, 2018.

[84] "A stochastic multi-criteria divisive hierarchical clustering algorithm," A. Ishizaka, B. Lokman, and M. Tasiou, Omega, vol. 103, p. 102370, 2021. [Online]. ScientificDirect.com/science/article/pii/S0305048320307246 is the link that can be accessed.

[85] "Community detection based on human social behavior," J. Sheng, J. Hu, Z. Sun, B. Wang, A. Ullah, K. Wang, and J. Zhang, Physica A: Statistical Mechanics and its Applications, vol. 531, p. 121765, 2019. [Online]. ScientificDirect.com/science/article/pii/S0378437119310246 is the link that can be accessed.

[86] J. Ma, J. Zhang, Y. Lin, and Z. Dai, "A multiobjective evolutionary algorithm reveals cost-efficiency trade-offs of the human brain network," NeuroImage, vol. 236, p. 118040, 2021. [Online]. https://www.sciencedirect.com/science/article/pii/S1053811921003177 is the URL that is accessible.

[87] In the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), A. Aylani and N. Goyal, "Community detection in social network based on useras social activities," pp. 625–628.

[88] "Afif: Automatically finding important features in community evolution prediction for dynamic social networks," by K. Kadkhoda Mohammadmosaferi and H. Naderi, appeared in Computer Communications, vol. 176, pp. 66–80, 2021. [Online]. https://www.sciencedirect.com/science/article/pii/S0140366421002176 is the URL that is accessible.

[89] An adaptation of the parliamentary metaheuristic for permutation constraint satisfaction was presented by L. de Marcos, A. Garc´ıa, E. Garc´ıa, J. J. Mart´ınez, J. A. Guti´errez,

R. Barchino, J. M. Guti´errez, J. R. Hilera, and S. Ot§on at the IEEE Congress on Evolutionary Computation, 2010, pp. 1–8.

[90] Q. Askari, I. Younas, and M. Saeed, Knowledge-Based Systems, vol. 195, p. 105709, 2020; "Political optimizer: A novel socio-inspired meta-heuristic for global optimization." [Online]. Accessible:/science/article/pii/S0950705120301350 available at https://www.sciencedirect.com

[91] Applied Soft Computing, vol. 87, p. 106010, 2020; E. Osaba, J. Del Ser, D. Camacho, M. N. Bilbao, and X.-S. Yang, "Community detection in networks using bio-inspired optimization: Latest developments, new results and perspectives with a selection of recent metaheuristics." [Online]. https://www.sciencedirect.com/science/article/pii/S1568494619307926 is the URL that is accessible.

[92] A novel evolutionary algorithm on communities detection in signed networks is presented by X. Zhu, Y. Ma, and Z. Liu in Physica A: Statistical Mechanics and its Applications, vol. 503, pp. 938–946, 2018. [Online]. https://www.sciencedirect.com/science/article/pii/S0378437118310550 is the URL that is accessible.

[93] Swarm and Evolutionary Computation, vol. 54, p. 100668, 2020; X. Wan, X. Zuo, and F. Song, "Solving dynamic overlapping community detection problem by a multiobjective evolutionary algorithm based on decomposition." [Online]. ScienceDirect.com/science/article/pii/S2210650218310563 is the link where you can find it.

[94] "The network data repository with interactive graph analytics and visualization," by R. A. Rossi and N. K. Ahmed, appeared in AAAI in 2015. [Online]. http://networkrepository.com is accessible .

Publication

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

Thorough Examination of Function-Based Community Detection via Social Media for Single and Multiple Purposes

Ravindra Singh Yadav[1],Dr. Balveer Singh[2]

[1]PhD Research Scholar, Department of Computer Science,P.K.University, Shivpuri, India
[2]Professor, Department of Computer Science, P.K.University, Shivpuril, India

Abstract –

As the Internet has grown, social networks have drawn interest as research subjects from a wide range of academic fields. The most precise systems are represented in intricate networks. Complex networks are most commonly characterized by their community structure, in which the relationships between node groups are more intimately linked to one another than with the network as a whole.Finding the main clusters and community structures in complex networks, like web charts and biological networks, enables the discovery of organizational rules. The communities appear to overlap in general. One of the defining characteristics of social networks is overlap, which is the state of an individual being a member of multiple social groups. Overlapping community discovery has drawn a lot of interest recently in social network application domains.Numerous approaches utilizing various tools and techniques have been put forth to address the issue of overlapping community discovery.This paper presents a thorough analysis of single and multi-purpose functions for community detection, as well as a comparative analysis of the heuristic overlap community detection algorithm over social media.

Keywords: single function, multi-purpose function, optimization algorithm, heuristic classification, community detection, and seed community

I. OVERVIEW

The 21st century has seen a rapid advancement in communication and technology, making information access necessary.and make the best use of it possible, making it an essential human need and a necessary component of their everyday existence. The Internet is without a doubt the most convenient and quick way to obtain information in the modern world. In addition to serving as a network that links millions of computers worldwide, the Internet also serves as a platform that links millions of individuals and thousands of social groups, and it is always expanding. One of the most widely used internet apps, social media is quickly rising to prominence as one of the most crucial communication tools available today. The rate at which people access social media rises in tandem with the frequency at which people use the Internet. Social media is predicted to play a near-essential role in internet usage in the near future. Social media apps aim to meet almost everyone's needs by employing technology that goes beyond simple communication. many subjects, including games, learning, and searching. People won't need another tool if they can find nearly anything they're looking for on social media. Along with Researchers can now access and analyze data on vast networks, such as social media, thanks to advancements in computer technology and network analysis.The analysis of individual and social group structures and behaviors in the press (separation, clustering, relationship determination), electronic commerce and online advertising (customer profile creation and trend analysis, personalised advertising and offering), physical structure analysis (transportation, installation, infrastructure), and analysis of large data sets (media tracking, academic publication analysis, genetic research) are just a few of the fields that use complex network analysis today [1]. Finding communities and communities in networks is the most pressing problem in

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

network analysis today. Identifying their communities in networks is used in a variety of fields, including engineering, physics, chemistry, biology, and social sciences. For instance, functional units of proteins can be identified or their functions can be predicted thanks to the discovery of biological communities [2]. Given the immunization interventions for infectious diseases in related networks and knowledge of the spread of viruses in social networks, community structure is a crucial topological feature in sociology [3].The ability to classify nodes based on the structures of the groups they belong to and to reveal groups is a crucial aspect of community discovery. In a social network structure, a set of nodes is referred to as a community if its link count exceeds the total number of connections outside of it. Communities, sometimes referred to as clusters or modules, are collections of nodes that work together in networks and typically share common functionalities [4]. Figure 1 provides a grid that schematically depicts the communities.The majority of community detection techniques are based on the division of links between groups. The most significant issue that arises in real-world network structures is the overlapping circumstance known as the potential for nodes to belong to multiple groups. Nonetheless, a lot of algorithms typically include nodes in a group because of the intricacy of the processes, disregarding overlap [5]. Accurate information about the structure of complex networks cannot be obtained by this grouping [6].For the purpose of identifying overlapping communities in intricate networks, numerous algorithms exist. The most popular algorithm is CPM. But CPM lacks the flexibility needed for actual networks. When the network is extremely dense, CPM detects meaningful clicks; when the network is sparse, it does not.Therefore, the network's capabilities have a significant impact on CPM.A Genetic Algorithm (GA) is used by GA-Net + [7] to adopt overlapping communities. The process creates a line chart from a node chart. The line chart's nodes display the node chart's edges, and the node chart's edges display the relationships between their neighborhoods[8]. The line chart is then presented as an overview of the genetic algorithm, and in order to achieve fit, it is transformed into a node chart at each stage [9].Other popular research for community discovery includes overlapping ensemble detection in networks [11], network communities [10], and an algorithm for rapidly identifying overlapping ensembles [12].Optimisation algorithms represent an additional technique for community discovery within social networks.The process of finding the best answer to a problem is called optimization. Heuristic optimization algorithms, which are widely used in daily life, are the basis for meta-heuristic optimisation algorithms, a decision-making mechanism [13]. It is intuitive, for instance, to make decisions at crossroads and move from one location to another based only on a feeling of direction, not knowing where the path will lead. Meta-heuristic algorithms are the structure that determines which approaches to use when three heuristic algorithms are beneficial for a problem from different perspectives.An overview of a heuristic community detection algorithm over social media is provided in this paper. The remainder of the document is structured as follows: The social media community and algorithm are described in Section 2. Single and multi-purpose function based metaheuristics community detection methods are presented in Section 3. A comparative analysis of the community detection algorithm over six different data sets is covered in Section 4. The paper is concluded in Section 5 with an outline of the foundational and ongoing work.

II. Social Network Community

Networks are used to represent the majority of complex networks.

The World Wide Web (WWW), for instance, is a network of linked webpages; social networks are networks in which individuals are represented as nodes and the relationships that bind them as edges. In a

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

similar vein, biological networks are made up of nodes that express biochemical molecules and boundaries that specify the connections among them [14]. The majority of research in recent years has been devoted to comprehending how network topology affects system dynamics, behavior, and network structure and evolution.Understanding complicated network architectures also requires identifying community structures [15].Groups of nodes are referred to as communities in networks.where ties inside groups are abundant and connections across groupings are rare. The association of people who communicate regularly is another definition of a community. Because of this, communities are essentially collections of nodes that interact similarly and have shared traits [16].Communities inside network architectures provide us specific information about people's study topics, interests, patterns, and other characteristics. In actual networks, theThe structure of networks is not uniform. Systems that focus and group together in a certain region, which we refer to as ensembles, are most likely collections of nodes with related features and functions [4].There are several concrete application areas in communities. For instance, by allocating the same servers to each client cluster, clustering web clients with comparable interests or locations together improves service performance on the WWW. In online buying systems, an efficient advice system between the buyer and the supplier may be built by identifying the community of consumers with similar interests[17].

By recognizing communities, hierarchical organizations in complicated real-world networks may be planned. Communities within communities are a common feature in real networks.The best example of a hierarchical arrangement is the human body.The body is comprised of organs, organ-derived tissues, and tissues composed of cells. Business firms are another instance of a hierarchical organization. One way to conceptualize business enterprises made up of midlevel workgroups is as a pyramid that extends from the workers to the top of the organization. Another meaning of "vertex similarity" is the concept that distance between nodes on a spaceplane serves as a similarity criteria. Traditional grouping techniques frequently employ this strategy. In case nodes are unable to be positioned on a spaceplane, an adjacency matrix may be employed. Even if they are not neighbors, it may be said that they are similar if their neighbors are also similar. Measuring the number of distinct pathways between two nodes, the length of the shortest route, or the random walk may also be used to identify commonalities between nodes [18].Although the initial research on identifying community structures implied that a node could be a part of only one community, networks are made up of various connections in which nodes can be found.be a part of multiple communities; this arrangement is known as overlap. Examples of human relations relationships between two people include those involving family, friends, and coworkers. Hence, one of the most important problems in the analysis of actual social networks is finding overlapping communities. In Figure 2, a network of three distinct communities is displayed. Four nodes in the network are part of multiple communities, demonstrating how communities in networks overlap.Since the community and modular structure are used to determine how well the systems function, they are regarded as crucial components of real-world social networks. Still, a lot of Effective and efficient community discovery techniques have been developed in response to uncertainties regarding community identification.

A. Conventional Approaches

a) Partitioning Graphs

It involves splitting the nodes into k groups of fixed sizes so that there are as few edges as possible between the groups. However, it is not a suitable approach for social network analysis when the number of groups present in social network structures is unknown beforehand. Iterative bi-sectioning is one of its

ISSN PRINT 2319 1775 Online 2320 7876

most important algorithms [19]. Min-Cut, Max-Flow Theorem. A dividing line where there are the fewest edges between groups is depicted in Figure 3 for k = 2.

b) Organizational Structure

Groups within social networks are typically entwined in a hierarchical structure. It is an approach that combines removing low-affinity nodes, grouping similar nodes, and dividing groups. Depending on the similarity criterion that is chosen, different results will be obtained [4].

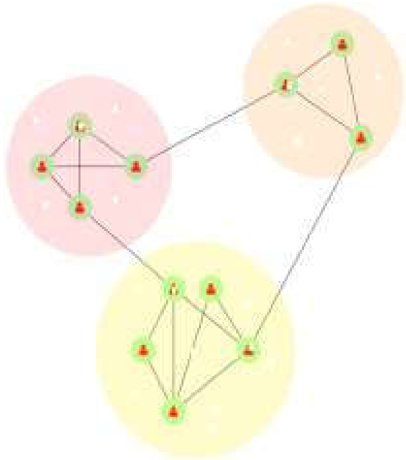


Figure 1 Network configuration with three communities

c) Dividing Grouping

Each node is viewed as a point in space, and the group number k is predetermined in this case. Assuming a given function, the objective is to partition the points into k groups based on how far apart they are from the center. The most often utilized components are k-means, k-central, k-median, and minimum k-clustering [20]. The drawback remains the same in this instance: knowing how many d) Clustering Spectrally Numerous strategies and procedures that divide data into sets using an eigenvector like S or another matrices generated by it [21]. This method involves taking the similarity matrix's eigenvectors and grouping them using a function like k-means [20].The Laplace matrix is the most often utilized matrix. This method allows the number of groups in the line to be determined from the eigenvector components.

B. Algorithms for Segmentation

It is a technique that seeks to identify and eliminate the edges in a graph that link groups in order to discriminate and reveal the teams. How to identify the edges that join these groups is crucial. The Girvan-Newman algorithm is the most widely used algorithm [22]. In this case, edges are chosen according to a factor known as edge centrality. Every edge's centrality value is computed. Deleted are the edges with the highest centrality value. Repetition of the first step and deletion of the edge with the highest value are the

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

next steps in the process.Edge betweenness, random walk edge betweenness, and current flow betweenness are also employed in addition to the edge centrality criteria [4].

C. Methods Based on Modularity

The most popular and frequently applied quality function in graph analysis is modularity. Though not entirely validated, a high Good groups are thought to be indicated by a modularity value [4]. A line is deemed to have a group structure if its modularity value is greater than that of a random line of the same size and degree on the graph. A high modularity value, however, does not always imply the existence of a group structure. Certain random graphs can have high modularity values even though they lack group structure.

There is no linear time solution to the NP-Complete problem of improving the modularity function. Nonetheless, successful algorithms with a range of convergences have been created [23, 24]. The alteration that maximizes the function of quality

D. Adaptive Formulas

The random walker model is the most widely used dynamic algorithm for community exploration.This technique results in the random walker remaining in the community for an extended period of time if the graph's connections have a high density; logically, this means that the chart has strong communities [4].

E. Alternative Approaches

In addition to the previously listed and widely used techniques, there are techniques based on statistical inference (Bayes, etc.) [25, 26], techniques that tag nodes and use the tag that their neighbors share the most in each iteration to split groups in this manner [4], click filtering techniques [27], approaches to deal with overlap, and multi-resolution techniques.

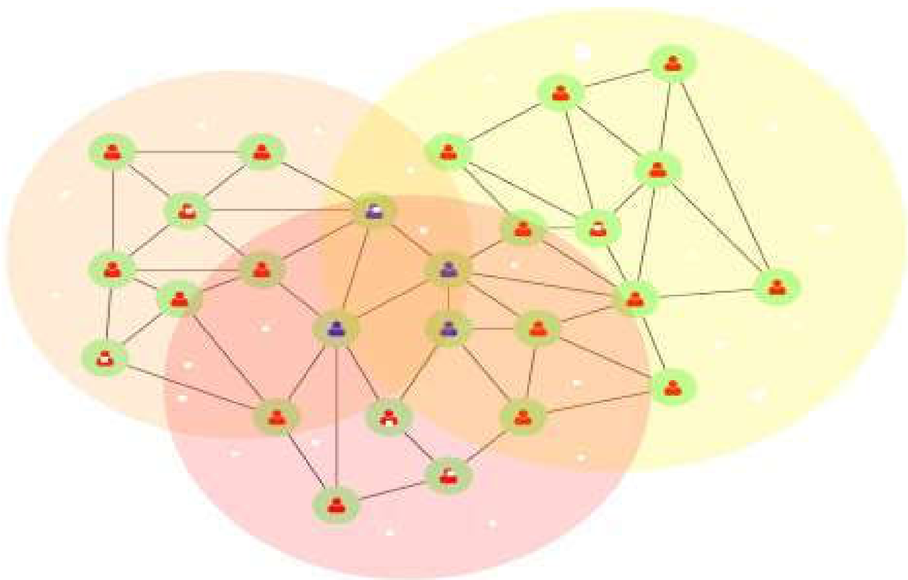


Fig. 2 Overlapping community

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

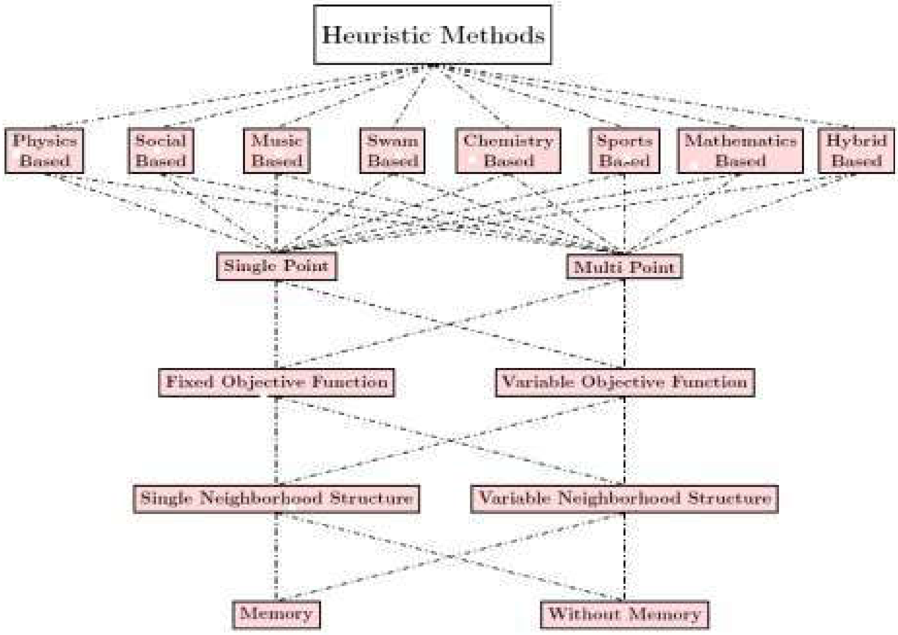


Fig. 4 Meta-Heuristic Methods

III. METHODOLOGICAL OPTIMIZATION METHODS Any issue pertaining to determining values for unknown parameters

An optimization problem is one that must be satisfied under specific constraints. Optimization entails optimization. Finding the best solution out of all the options for a problem under a certain set of circumstances is the task at hand. In other words, heuristic algorithms have the ability to converge but do not ensure a precise result. This circumstance offers a resolution that is nearly perfect [29].

Heuristic algorithms are necessary for the following reasons:

• It's possible that the optimization problem has a structure that makes it impossible to pinpoint the precise solution.

• Heuristic algorithms have the potential to be much more clear easier to make decisions for.

• Heuristic algorithms are useful for learning and for assisting in the precise solution's discovery.

• The most difficult parts of real-world problems (such as which objectives and constraints should be applied, which alternatives should be tested, and how to gather problem data) are frequently overlooked in definitions created using mathematical formulas. More substantial errors could result from inaccurate data used during the model parameter determination stage than from the heuristic approach's potential for producing a suboptimal solution [29].

A decision-making process based on heuristic optimization algorithms—which are widely applied in daily life—is known as a meta-heuristic optimization algorithm [30].

IJFANS International Journal of Food And Nutritional Sciences

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

The meta-heuristic Algorithms, which are becoming more powerful and well-liked in recent years, solve search or optimization problems using a straightforward methodology. The following is a summary of the cause of these:

a. They also provide general approaches to solving the problem that can be used when various kinds of objective functions, constraints, and decision variables are present. The type of objective function, limiters, and variables used in the problem modeling are not relevant factors in determining the strategies for solving the problem.

b. It is independent of delimiters, decision variable count, and type of solution space.

c. There is no need for highly defined mathematical

models that can occasionally not be used due to the high cost of solution time and are challenging to set up for the model and purpose function of the system.

d. They don't require a lot of computing time because they have strong processing power.

e. It is simple to change and adjust to them.

f. It produces good outcomes for large-scale nonlinear and multinational problems.

g. Unlike classical algorithms, a solution algorithm for a given problem does not necessitate some assumptions that may be challenging to verify in adaptation. h. Similar to classical algorithms, it doesn't call for modifications to the target problem. They modify themselves to address various problems.These benefits make meta-heuristic algorithms widely used in a variety of fields, including computers, engineering, and man-made intelligence, and new versions are advised.

Figure 4 illustrates general-purpose meta-heuristic methods. They are bio-based (evolutionary algorithms, ant colony algorithms, bee colony algorithms, artificial immune algorithms, firefly algorithms, enzyme algorithms, sapling development algorithms, invasive weed optimization, monkey search algorithms, bacterial bait search algorithms), physics-based (multi-point heat treatment algorithms, electromagnetism algorithms, particle collision algorithms, big bang big crash algorithms), swarm-based (particle swarm optimization, ant colony optimization, bee colony optimization), and social-based (multipoint taboo Eight distinct approaches: investigation algorithm, parliamentary optimization algorithm, imperialist competitor algorithm), sports-based (league championship algorithm), music-based (harmony search), and chemistry-based

The group evaluates techniques (such as artificial chemical reaction optimization algorithms) and mathematics-based techniques (such as meta-heuristics and base algorithms)[31]. Additionally, hybrid approaches that combine them exist.Even though there are many highly effective algorithms and techniques in the literature, it is crucial to design, develop, and implement new strategies in the scientific field with the continuous improvement philosophy and always strive for better. Additionally, new meta-heuristic algorithms are continuously being proposed because the algorithm that yields the best results for all problems has not yet been designed. It is suggested that the current ones operate more efficiently.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

Because of this awareness, scholars have successfully added new meta-heuristic techniques to the body of literature in recent years. A. Algorithms for Socially Based Meta-heuristic Optimization

The literature contains a large number of recently suggested social-based heuristic optimization techniques. The most renowned The most commonly utilized among them is the Tabu search algorithm. Others have been filed more recently [32].

a) Algorithm of Imperialist Competitor

The Imperialist Competitor process (ICA) starts the process by generating an initial population, just as analogous evolutionary algorithms. The top few nations in terms of initial population are selected as imperialist nations, and the remainder people become imperialist colonies. The states that make up the empire divide up all of the specified regions. The colonies start to gravitate in the direction of the right imperialists after being distributed among the imperialist powers. Empires' strength is determined by the imperialist's capacity and their their areas ceded to the imperialists. The algorithmic process is ongoing, with the race between the imperialists having begun. Not able to

The imperialism will be removed from the race if it manages to get stronger or achieve success. Robust empires gain power during the race, whereas weak kingdoms deteriorate and head toward collapse. The race goes on until there is only one empire left, at which point other nations become colonies of the empire that survives thanks to an algorithm. Territories and imperialists will hold equal status and authority in the utopian society that is created at the conclusion of the race [33].The algorithm's flow chart is displayed in Figure 5.

b) Educating Algorithms for Learning-Based Optimization

The Teaching Learning Based Optimization is another newly created meta-heuristic optimization technique. optimization issue, with the fitness function's best value being the optimal solution. There are two scenarios in which the TLBO algorithm operates: The Method of Instruction and The Process of Learning[36].It is well acknowledged that the instructor plays a crucial role in the teaching process as the one who imparts information to the students. Students are a clear indicator of a teacher's quality. It has been shown that when pupils have excellent professors, both their circumstances and grades improve.

As a result, the interaction between the teacher and the student affects the teaching process. Students are the primary component in the learning process [37].To help in understanding the phases of the TLBO algorithm, a flow chart has been produced, as shown in Figure 6.

Algorithm (TLBO) [34]. The TLBO algorithm operates based on how a teacher affects their pupils in a classroom. The teaching and learning capacities of instructors and pupils at a school are described by the algorithm. Two fundamental elements of this algorithm are the teacher and the pupil [35].The population in the algorithm is the group of students, and other design variables for the optimization issue are the many disciplines that are taught to the pupils. The outcome of a student is comparable to the optimization problem's fit value. For the entire population, the teacher is thought to be the best option.The vocabulary employed in design

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

c) Algorithm for Social-Emotional Optimization

A novel social-based optimization method called the Social-Emotional Optimization Algorithm (SEOA) mimics human conduct [38]. The human community is linked to the term "social." The community's residents work to elevate their social standing.

The SEOA's operational processes are listed in the algorithm.1. A step-by-step breakdown of the algorithm for social-emotional optimization

1. Get going

2. Every person is created one after the other, and the problem space is randomly assigned to each one of their starting places.

3. The goal function is used to compute each person's fitness value.

4. j. Based on the person's emotional index, their behavioral actions are dictated.

5. The whole population's location is updated.

6. The emotional quotient is calculated.

The optimal option, if the termination condition is satisfied,

7. The best option is approved if the termination requirement is satisfied. Step 2 is returned if the condition is not satisfied.

8. Come to an end.

Every individual in SEOA is a virtual person. People base their decisions on how to behave at each phase on the corresponding emotional index [39]. There are three categories for the emotional index: low, medium, and high. An action is chosen based on the emotional index. Depending on whether the intended behavior is accurate, the status value is recycled from society in accordance with the chosen behavior. The person's emotional index rises if this decision raises the social status value. If not, the social status value drops as the emotional index decreases [40].

c) Organizing Brainstorms

In generally acknowledged organizations, brainstorming is a popular technique to foster creativity, such as fostering creative contemplating. Osborn created brainstorming for the first time at the advertising agency in 1939. He organized this approach to problem-solving in Applied Imagination before the end of 1957 [41, 42]. Following then, brainstorming sparked a global interest in academics and business. People from various ethnic backgrounds get together during the brainstorming process to collaborate and interact in order to provide brilliant ideas for solving problems. The phases in the BFOA process that were created using brainstorming inspiration are listed in Algorithm Two steps to the Brain Storming Optimization Algorithm explanation:

ISSN PRINT 2319 1775 Online 2320 7876

1. Get going

2. Individuals who might be possible solutions are produced.

3. There are n individuals and m groups.

4. N people are assessed.

5. Each cluster's members are ranked, and the cluster's center is assigned to the best member.

6. A value in the range of 0 to 1 is created at random.

(a). In the event that a generated value is smaller than P5a's predefined value

(i). At random, select a cluster center.

(ii). Create a random person to take the place of the selected cluster center.

7. Generate new persons.

(a). It generates a random number between 0 and 1.

(b). Should the generated value fall short of P6b,

6 (i). Select a random set a with probability P

(ii). Create a random number in the range between 0 and 1.

(iii). If the amount is below the predetermined

value of P6b

iii, 1) To create new people, choose the cluster center and add a random value.

If not, select a random person from the cluster and combine them with the value that was created at random to create new people.

(c). If not, choose two clusters at random to produce new people.

(i). Create a value at random.

(ii). Choose and combine two cluster centers, then add the randomly produced value to create additional people if the generated value is smaller than the predefined probability of P6c.

(iii). If not, two individuals are chosen at random from each cluster to merge, and the created value is added to create new individuals.

8. If n additional people are created, proceed to step 9 and then step 7.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

9. The point at which the predefined maximum number of iterations has been achieved; if not, proceed to step 3.

10. Stop.

e) Algorithm for Group Leaders Optimization

The Group Leaders Optimization Algorithm (GLOA) is an algorithm that was created through evolution, drawing inspiration from the impact of social group leaders. The issue space is split up into various groups, and a leader is chosen for each group [43].Each group's members don't have to be similar characters; they can be chosen at random. Each group selects its best member to be the leader. Every iteration, members of each group want to look like their leaders. The method establishes a solution space between the group members and the leader in this manner. Following a few actions, it was noted that group members had a leaderlike appearance. One person is picked at random to promote variety within the group. A few of The variables of the other group members take their position. Furthermore, a crossover operator facilitates the group's arrival to the local minimum, and another search of the solution space can be conducted to boost variety [44]. Figure 7 f) outlines the algorithmic procedures by which n groups of P members are created and group leaders are chosen based on their appropriateness values. An Algorithm for Social Hierarchies

The social behaviors seen in a range of biological systems and human organizations served as the model for the Hierarchical Social Algorithm (HSA). Several problems with infinite resources have been successfully solved using this meta-heuristic technique. The concurrent optimization of the set of appropriate solutions is the fundamental principle of HSA. Every social group has a workable answer, and these groupings are first dispersed at random to create distinct locations for the solutions. Each group uses development tactics to compete with their neighbors or improve their target function. In this instance, relevant social rivalry and collaboration yield a superior outcome 20. The objective solution is therefore optimized. The optimal solution is identified in a single group at the end of the procedure [1].

g) Algorithm for Human Group Formation

The Human Group Formation Algorithm (HGFA) is a modern social-based meta-heuristic optimization algorithm that draws inspiration from the behaviors of both outgroup and ingroup members that make an effort to stay as close to their groups as feasible. Sociologists have established what constitutes an in-group and out-group in order to categorize people into social groups. Those who are accepted by the group as members of the in-group include which they are a part of. When someone is classified as a member of a group, they identify with that group and believe that they are unique from other groups. They believe that their group is superior to all others. Because of this, even while they are apart from the group, members of the group make every effort to keep the group together [47]. It demonstrates the conversion of the ideas presented in Figure 8 into applications.

h) Algorithm With Social Basis

A novel algorithm known as a Social Based Algorithm (SBA) blends an evolutionary algorithm with a socio-political an Imperialist Competitor Algorithm-based procedure.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

People reside in a variety of communities, including multinational, republican, autocratic, and monarchical ones. Additionally, each community has a unique style of leadership. This strategy aims to include a small number of individuals in the community development trait [48]. Algorithm 3 displays the SBA's process phases.

Algorithm 3: The Social Based Algorithm explained step-by-step:

1. Launching

2. Filling up the parameters

3. adhere to

(a). Clearly stating the optimization issue,

(b). Creating a random population

(c). Choosing a few powerful individuals at random to serve as leaders,

(d). Placing the remaining people at random throughout various areas,

(e). Using the imperialist cost function to launch empires T.Pci, f. the selection of charismatic leaders to be emperors,

4. Ten circles Nd is equal to Nd+1.

5. where i = 1, 2,..., N

(a). Choice (b). Cross (c). Change (d). Substitution

6. Let i be 1, 2,..., N

(a). The leaders of each group are relocated to their kingdom as part of the assimilation program of humans

(i). $x \sim U$ (0, absorption inside x)

d) (ii). $d:$ The separator and imperialist's distance

(b) The revolution of the people

(c). Countries adopt an assimilation program in which the leaders of each group establish their empire and the populace of each nation follows suit

(i). $x \sim U$ (0, external assimilation coefficient x d)

(ii). $d:$ The separation between the imperialist and the leader

(d). The nation-state revolution

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

(e). Modifying the address

(f). imperialist race: selecting the weak nation from the weak empire and awarding it to the empire with the highest probability of possessing it

(g). Elimination; the abolition of empire and the powerless principle

7. Verifying the termination requirement and continuing from steps 4 through 7 until it is satisfied.

8. Stop.

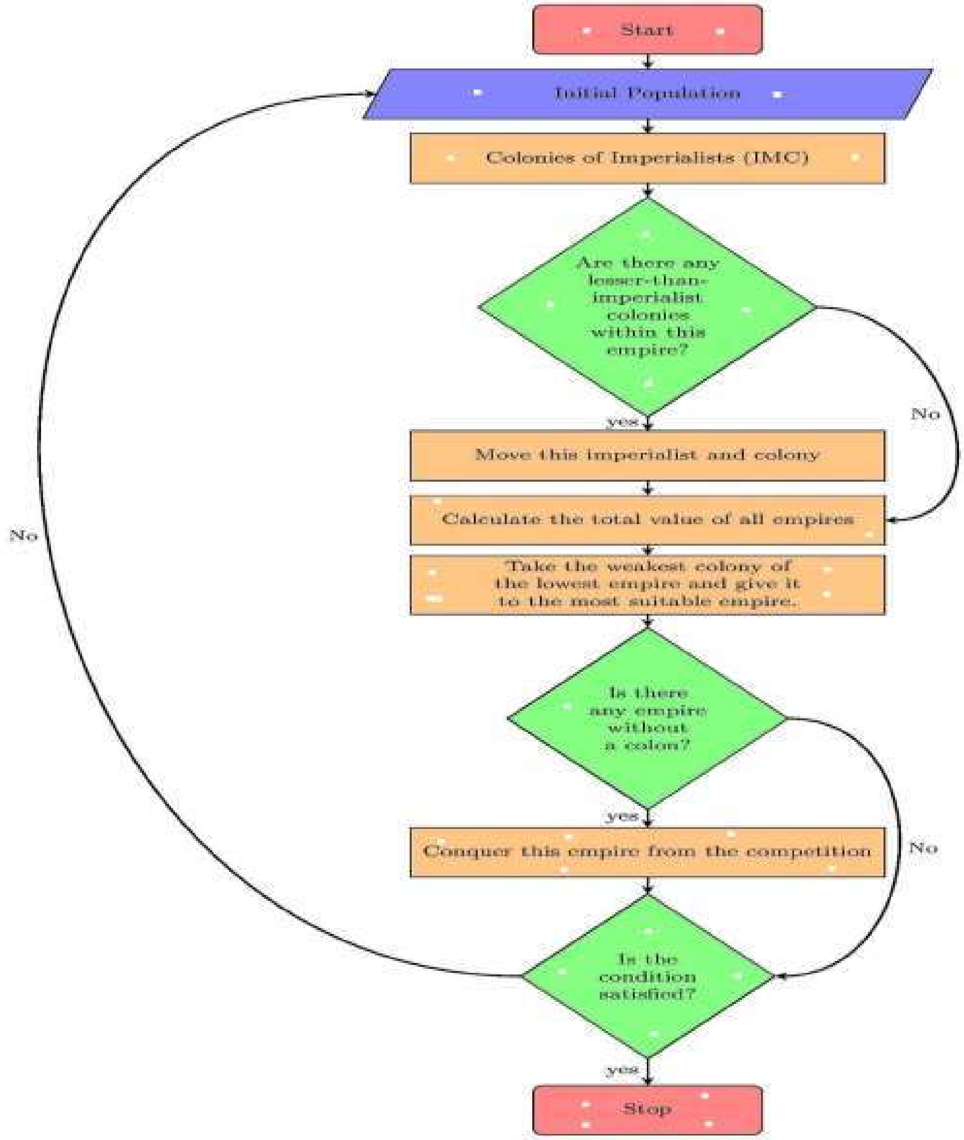


Fig. 5 Flow Chart Of Imperialist Competitive Algorithm

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

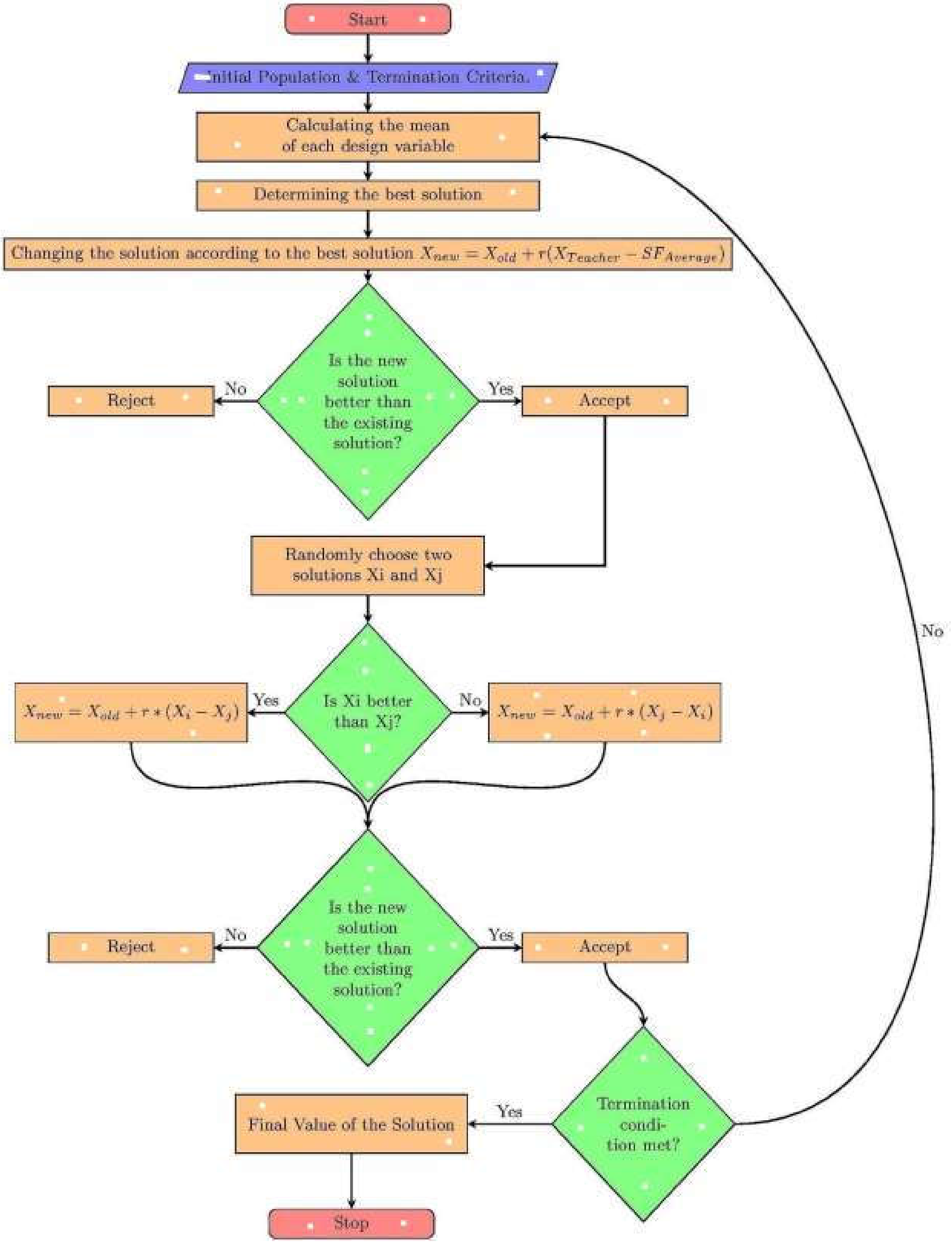


Fig. 6 Flow Chart Of TLBO Algorithm

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

IV. BENCHMARK COMMUNITY DETECTION ALGORITHM SOCIAL THEORY RESULT ANALYSIS

Six distinct graphical social media data sets—Word adjacencies, Zachary Karate Club [49], Dolphin social network [50], Les Misérables, books about US politics, and American college football [51]—were used in the performance evaluation to determine the effects of single and multi-purpose based heuristic community detection algorithms. The evaluation parameters included modularity and normalized mutual information.

A structural evaluation of networks called modularity assesses how well a subgraph—a group, cluster, or community—in the network can be used to extract community structure [52]. Communities in a particular network develop as a result of groupings of nodes with higher modularity being relatively dense with one another in the network:

HA = 1 2|HA

$$M = \frac{1}{2|E|}\sum_{\text{xy}}\left[e_{xy} - \frac{w_x w_y}{2|E|}\right]\delta(C_x, C_y)$$

$$= \sum_{i=1}^{n} fii - f'^2_i$$

A probabilistic function (cx,cy) equals 1 if both nodes x and y belong to the same community structure, and 0 otherwise. fii represents the edge in the community i, and F′ is the belonging probability of a random edge to the community i that is attached to vertices in the community i. Where exy represents the edge from node x to node y, Wx represents the summation of the weights of the edges linked to node x, and cx is the belonging community structure of node x. On the other hand, normalized mutual information is a way to scale the similarity between intra community nodes by normalizing the intra-community mutual information score:

The 0 node and nmi(x, c) are completely different.

One node is identical to another

Mutual information may be computed using the formula

$$\text{nmi}(\text{x}, \text{c}) == \frac{2 * \text{i}(x, ci)}{e(x) + e(c)}$$

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

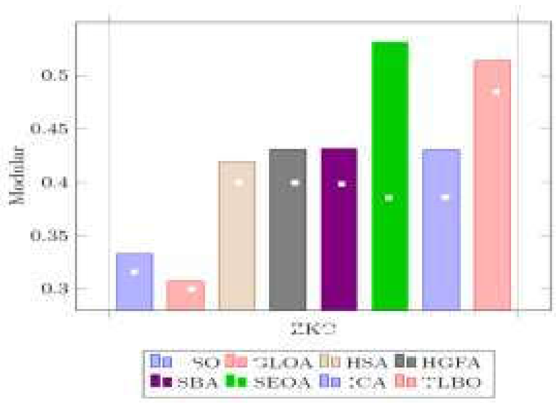
Table 1: A Comparative Examination of Social Theory's Effect on Modularity

Method of Classification	Modularity					
	DCN	BUP	LM	WA	ZKC	ACF
ICA	0.6179	0.6714	0.6122	0.4192	0.4302	0.7211
SEOA	0.661	0.61	0.6121	0.4118	0.5313	0.6107
BSO	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
GLOA	0.6179	0.6714	0.6122	0.4192	0.4302	0.7211
HSA	0.661	0.61	0.6121	0.4118	0.5313	0.6107
SBA	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
HGFA	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
TLBO	0.6196	0.7119	0.6217	0.5103	0.5137	0.6204

Table 2: Comparison of Social Theory's Effects on Normalized Mutual Information

Method of Classification	Mutual Information Normalized					
	DCN	BUP	LM	WA	ZKC	ACF
ICA	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
SEOA	0.6179	0.6714	0.6122	0.4192	0.4302	0.7211
BSO	0.661	0.61	0.6121	0.4118	0.5313	0.6107
GLOA	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
HSA	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213
SBA	0.6179	0.6714	0.6122	0.4192	0.4302	0.7211
HGFA	0.661	0.61	0.6121	0.4118	0.5313	0.6107
TLBO	0.4844	0.5674	0.5105	0.3611	0.3326	0.6213

The class label performance evaluation of the benchmark community detection algorithm with and without social theories is displayed in tables 1 and 2 as modularity and normalized mutual information, respectively, where c is the community structure, e is the entropy, and i(x;c) is the information gain for element ci. The integration of social theories with the community detection algorithm leads to a considerable improvement in both assessment parameters. The approximate results of the community detection algorithms



BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO 62.11%, 86.18%, 71.53%, 82.43%, 81.02%, 86.51%, and 86.24% NMI over 33.26%, 30.76%, 41.95%, 43.05%, 43.14%, 53.13%, 43.02%, and 51.37% modularity

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

ZKC datasets, as seen in Figures 9 and 10.

The highest NMI information is achieved by the ICA and TLBO algorithms, but the SEOA method leads in modularity.

Fig. 10 Normalized Mutual Information for ZKC Data Set Community Detection

Fig. 11: Community Detection's Modularity Across AFC Data Set

On the other hand, as illustrated in figures 11 and 12, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain approximately 62.13%, 59.47%, 59.29%, 72.03%, 60.14%, 61.07%, 72.11%, and 62.04% modularity over the AFC dataset. While SEOA and HAS algorithms accomplish modularity, ICA algorithm leads the way.

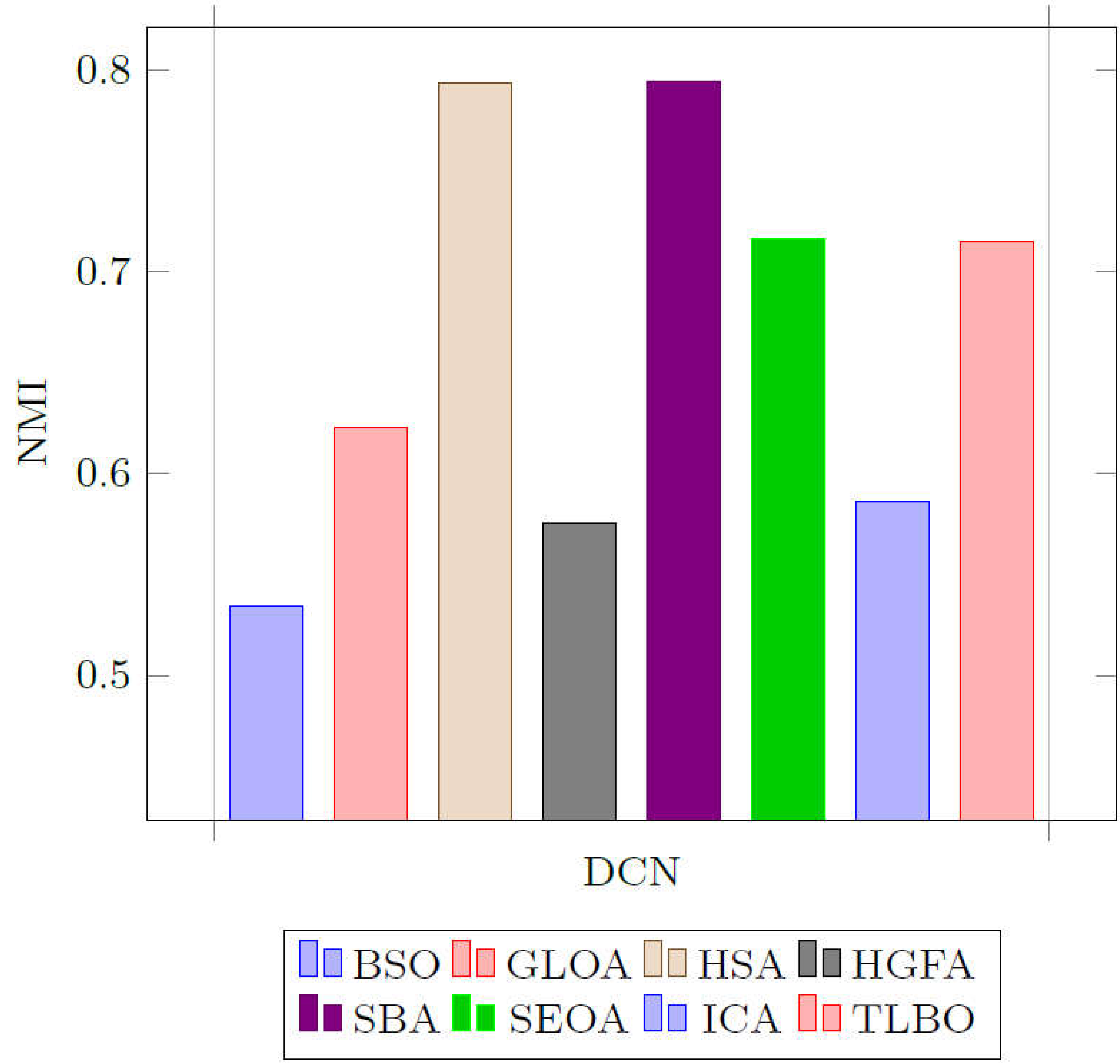
the highest NMI data. In contrast, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain around 56.74%, 51.75%, 56.01%, 67.68%, and 61.57% over the BUP dataset.

As seen in figures 15 and 16, there is 61.51%, 67.14%, 71.19% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, and 52.51% NMI, respectively. The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. On the other hand, as illustrated in figures 17 and 18, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO on the LM dataset gain approximately 51.05%, 50.98%, 58.61%, 60.79%, 60.71%, 61.21%, 62.17% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, and 52.51% NMI, respectively. The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. In contrast, the WA dataset, BSO, GLOA, and HSA community discovery algorithms About 56.74%, 51.75%, 56.01%, 67.68%, 61.57%, 61.51%, 67.14%, and 71.19% modularity are gai

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022



ned by BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO, whereas 42.25%, 52.09%,

Figures 15 and 16 illustrate 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, and 52.51% NMI, respectively. The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. On the other hand, as illustrated in figures 17 and 18, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO on the LM dataset gain approximately 51.05%, 50.98%, 58.61%, 60.79%, 60.71%, 61.21%, 62.17% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, and 52.51% NMI, respectively. The maximum NMI information is achieved by the SEOA algorithm, whereas the TLBO method leads in modularity. Conversely, the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO gain almost the same across the WA dataset.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

about 39.15%, 39.27%, 32.84%, 45.82%, 43.01%, 58.22%, 42.68%, and 51.22% NMI and 36.11%, 34.92%, 30.30%, 35.32%, 41.87%, 41.18%, 41.92%, and 51.03% modularity

as indicated by figures 19 and 20. While the SEOA algorithm produces the maximum NMI information, the TLBO algorithm wins in modularity.greater ACF network density, higher performance rate, and comparatively lower over the less dense WA dataset. Heuristic overlap across six distinct social media-based datasets for community recognition algorithms. According to this study, the performance rate of the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO varies with network density over social media datasets. It obtains a greater performance rate in dense ACF networks and a significantly lower one in weakly packed WA datasets. Additionally, across the greater dense network, TLBO and ICA are able to extract a higher informative community.

Concurrently, SEOA and the less dense networks. Network density affects how well community discovery algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO perform on social media datasets. It accomplishes a

greater ACF network density, higher performance rate, and comparatively lower over the less dense WA dataset. Heuristic overlap across six distinct social media-based datasets for community recognition algorithms. According to this study, the performance rate of the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO varies with network density over social media datasets. It obtains a greater performance rate in dense ACF networks and a significantly lower one in weakly packed WA datasets. Additionally, across the greater dense network, TLBO and ICA are able to extract a higher informative community.In addition, SEOA and HAS need to achieve greater outcomes with the less dense

V. CONCLUSION AND FUTURE WORK

In order to provide a thorough analysis of overlapping community structure on social networks, this research is regularly faced in daily life, and resolves community discovery that aligns with an approach that has never been used previously. The investigation and analysis have shown that the methods created for identifying overlapping communities in social networks offer answers to this issue by focusing on a single goal. This research also provides a comparative examination of six distinct social media-based data sets using meta-heuristic overlapping community recognition techniques.This study found that the performance rates of the community detection algorithms BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, and TLBO varied with the density of the social media data set. It achieved greater performance rates in the dense ACF network and significantly lower ones in the minimally packed WA data set. Additionally, across the greater dense network, TLBO and ICA are able to extract a higher informative community. Simultaneously, the less dense networks produce superior outcomes for SEOA and HSA.

[1] Social networks using hierarchical community identification, in Proceedings of the 7th International Conference on Information and Knowledge Technology (IKT), 2015 1–5.

[2] Communityguided link prediction in multiplex networks: Karimi, S. Lotfi, and H. Izadkhah, Journal of Informetrics, 15(4) (2021) 101178.

ISSN PRINT 2319 1775 Online 2320 7876

[3] A. Fang, Applied Mathematics and Computation, 406 (2021) 126262, The effect of communication structure on opinion dynamics in social networks with various true states.

[4] W. Liu, G. Hu, L. Chen, and T. Suzumura, 2017 IEEE International Conference on Big Data (Big Data), (2017) 3342–3351. A generalized incremental bottom-up community discovery approach for highly dynamic graphs.

[5] In the 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), (2019) 513–516, Q. Chen and L. Wei, Overlapping community detection incomplex networks: A survey.

[6] N. C. Debnath, P. Choudhury, R. A. Haraty, and D. K. Singh 2020 IEEE International Conference on Industrial Engineering: A study of dynamic community identification techniques in complicated networksTechnology (ICIT), 989–994 (2020).

[7] In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 1–3, S. Liu and Z. Li, A modified genetic algorithm for community discovery in complex networks.

[8]A novel evolutionary approach for overlapping community discovery was presented by B. Shen, N. Wang, and H. Qiu at the Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing in 2014, pages 766–769.

[9] Adaptive community discovery in complex networks using genetic algorithms, M. Guerrero, F. G. Montoya, R. BaAos, A. Alcayde, and C. Gil, Neurocomputing, 266 (2017) 101–113.

[10] Y. Jejani, T. Patalia, D. Joshi, and K. Fatania, An analysis comparing overlapping community detection in a large, intricate network, in Recent Advances in Energy-efficient ICRAECC (Computing and Communication), 2019; 1–6.

[11] Similarity preserving overlapping community discovery in signed networks, C. He, H. Liu, Y. Tang, S. Liu, X. Fei, Q. Cheng, and H. Li, Future Generation Computer Systems, 116 (2021) 275–290.The article "Detection of hierarchical overlapping communities in complex networks" was published in the 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS) and was written by Y. Yan and H. Xie.

[13] Performance comparison of physics-based meta-heuristic optimization algorithms by D. DEMIROL, F. OZTEMIZ, and A. KARCI, in International Conference on Artificial Intelligence and Data Processing (IDAP), (2018) 1–5.In Online Social Networks and Media, 3(4) (2017), 32–48, G. Rossetti, D. Pedreschi, and F. Giannotti, Nodecentric community discovery: From static to dynamic social network analysis.

[15] Pythia, a system for social media post topic discovery online, was developed by I. Litou and V. Kalogeraki. It was presented at the IEEE 37th International Conference on Distributed Computing Systems (ICDCS) in 2017, pages 2497–2500.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

[16] V. Moscato and G. SperlA, Knowledge-Based Systems, 224 (2021) 107112, A survey about community detection over online social and heterogeneous information networks

[17] M. Li, L. Wen, and F. Chen, Physica A, Statistical Mechanics and its Applications, 561 (2021) 125140, A novel collaborative filtering recommendation approach based on soft coclustering.

[18] Social set analysis: A set-theoretical approach to big data analytics, R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, IEEE Access, 4 (2016), 2542–2571.

[19] Sphynx, A parallel multigpu graph partitioner for distributed-memory systems, S. Acer, E. G. Boman, C. A. Glusa, and S. Rajamanickam, Parallel Computing, (2021) 102769.

[20] Information Sciences, 569 (2021) 596–614, F. J. BaldA˜ ¡n and J. M. BenAtez, Multivariate time-series classification through an interpretable representation.

[21] Scalable spectral clustering for overlapping community detection in large-scale networks: H. Van Lierde, T. W. S. Chow, and G. Chen, IEEE Transactions on Knowledge and Data Engineering, 32(4) (2020), 54–767.

[22] H. Jain and G. Harit, Using community detection for unsupervised temporal segmentation of human action, in the 25th IEEE International Conference on Image Processing (ICIP), (1892–1896), 2018.

[23] Pattern Recognition Letters, 122 (2019), 14–22; N. F. Haq, M. Moradi, and Z. J. Wang, Community structure detection from networks with weighted modularity.

[24] A new technique for overlapping community detection using multiobjective optimization, M. Ebrahimi, M. R. Shahmoradi, Z. Heshmati, and M. Salehi, Physical A, Statistical Mechanics and its Applications, 505 (2018) 825–835.

[25] Overlapping community detection in weighted temporal text networks: R. Dong, J. Yang, and Y. Chen, IEEE Access, 8 (2020) 58118–58129.

[26] Limited spectral clustering for dynamic community detection by A. Karaaslanli and S. Aviyente, in ICASSP IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2020, 8474–8478.

[27] Wang, G., Liu, S., and X. Ji, W. Physica A, Statistical Mechanics and its Applications, 566 (2021) 125612, 2021. Percolation on interdependent networks with cliques and weak interdependence.

[28] A. Baig, J. Qadir, S. Latif, M. S. Younis, and M. A. Javed, A multidisciplinary review on community detection in networks was published in the Journal of Network and Computer Applications, 108 (2018), pages 87–111.

[29] Research on applying artificial intelligence algorithms in a directed graph by Y. Zhou and F. Xu was published in the 2017 International Conference on Computing Intelligence and Information System (CIIS), pp. 116–120.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

[30] Alexandria Engineering Journal, 60(2) (2021) 2193–2208, 2021; M. Ehteram, A. Ferdowsi, M. Faramarzpour, A. M. S. Al-Janabi, N. Al-Ansari, N. D. Bokde, and Z. M. Yaseen, Hybridization of artificial intelligence models with nature-inspired optimization algorithms for lake water level prediction and uncertainty analysis.

[31] T. Li, J. Yang, and D. Cui, A benchmark microgrid's performance optimization control through artificial intelligence-based algorithms in multi-access edge computing, Physical Communication, 44 (2021) 101240.

Computers and Structures, 244 (2021) 106395; Ficarella, E., Lamberti, L., and Degertekin, S., Comparison of three novel hybrid metaheuristic algorithms for structural optimization problems.

[33] Mashayekhi, Dehghani, and Sharifi, M. Applied Mathematical Modelling, 93 (2021) 1–27, presents an effective imperialist competitive algorithm with likelihood assimilation for optimizing the topology, shape, and sizing of truss structures.

[34] Y. Ma, J. Song, L. Chen, X. Zhang, Knowledge-Based Systems, 212 (2021) 106599, A modified teaching-learning-based optimization algorithm for solving an optimization problem.

In Expert Systems with Applications, 148 (2020) 113246, Y. Zhang and Z. Jin present a novel metaheuristic method for solving global optimization problems through group teaching of optimization algorithm.

[36] Teaching-learning based optimization algorithm for core reload pattern optimization of a research reactor, A. Ammar and S. ul Islam Ahmad, Annals of Nuclear Energy, 133 (2019), 169–177.

[37] Applied Mathematical Modelling, 77 (2020) 309–326; A. K. Shukla, P. Singh, and M. Vardhan. An adaptive inertia weight teaching–learning-based optimization algorithm and its applications.

[38] M. Kumar, A. J. Kulkarni, and S. C. Satapathy, An optimization methodology inspired by society: the social evolution learning algorithm, Future Generation Computer Systems, 81 (2018) 252-272.

[39] Zhang, Y., and Zhang, P.: Machine learning and parameter configurations for support vector machines using a social-emotional optimization algorithm, Pattern Recognition Letters, 54 (2015) 36–42.

[40] Social-emotional optimization algorithm-based identification of nonlinear Hammerstein model by P. S. Pal, S. Choudhury, A. Ghosh, S. Kumar, R. Kar, D. Mandal, and S. P. Ghoshal, in 2016 International Conference on Communication and Signal Processing (ICCSP), (2016) 1633– 1637.

[41] Modified brain storm optimization algorithm in objective space for a pollution-routing problem, A. Rauniyar, R. Nath, and P. K. Muhuri, 2019 IEEE Congress on Evolutionary Computation (CEC), (2019) 242-247.

An optimally identified framework with brain storm optimization for multimodal optimization problems is presented in [42] Z. Dai, W. Fang, K. Tang, and Q. Li, Swarm and Evolutionary Computation, 62 (2021) 100827.

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, 2022

[43] Opinion leader identification in online social networks using the whale optimization algorithm by L. Jain, R. Katarya, and S. Sachdeva, Expert Systems with Applications, 142 (2020) 113016.

[44] A novel intelligence algorithm based on the social group optimization behaviors, X. Feng, Y. Wang, H. Yu, and F. Luo, IEEE Transactions on Systems, Man, and Cybernetics, Systems, 48(1) (2018) 65–76.

[45] B. Lokman, M. Tasiou, and A. Ishizaka, Omega, vol. 103, p. 102370, 2021: A stochastic multi-criteria divisional hierarchical clustering algorithm.

[46] "Community detection based on human social behavior," J. Sheng, J. Hu, Z. Sun, B. Wang, A. Ullah, K. Wang, and J. Zhang, Physica A, Statistical Mechanics and its Applications, 531 (2019) 121765.

[47] Cost-efficiency trade-offs of the human brain network revealed by a multiobjective evolutionary algorithm, NeuroImage, 236 (2021) 118040, J. Ma, J. Zhang, Y. Lin, and Z. Dai.

[48] In 2017 International Conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 625–628. A. Aylani and N. Goyal, Community detection in social network-based non user as social activities.

[49] Zachary, W. W. Journal of Anthropological Research, 33(4) (1977) 452–473, presents an information flow model for conflict and fission in small groups.

In the Proceedings of the Royal Society of London, Series B, Biological Sciences, 270(2) (2003), S186–S188, D. Lusseau discusses the emergent properties of a dolphin social network.

Community structure in social and biological networks, M. Girvan and M. E. J. Newman, Proceedings of the National Academy of Sciences, 99 (2002) 7821–7826.

Modularity and community structure in networks: a study by M. E. J. Newman, Proceedings of the National Academy of Sciences, 103(23) (2006) 8577–8582.

Journal of Namibian Studies, 33 S1 (2023): 3826-3852 ISSN: 2197-5523 (online)

Multifunctional And Function-Oriented Parliamentary Streamlining Structure For Community Identification On Social Media

Ravindra Singh Yadav*,Dr. Balveer Singh

[1]Department of Computer Science, P.K. University Shivpuri ,Madhya Pradesh, India.

Abstract

This paper presents a multi-purpose function-based parliamentary optimization (MPPOA) community detection methods. Initially, the population of parliamentary optimization algorithm (POA) was created in a python environment for the data used. The population formed was divided into a certain number of groups, and the power values of each group were calculated. While strong groups show joining according to the determined combination probability value, the vulnerable groups are eliminated from the population according to the determined deletion probability. The result for the problem has been approached. The program steps continued until all groups were combined and the termination condition of the algorithm was met; the individual with the highest eligibility values among the remaining data in the last stage was accepted as the solution to the overlapping community discovery problem of the proposed algorithm. Subsequently, performs a proposed work evaluated over one artificial and four real network-based social media data sets. The comparison of feature evaluation is carried out to identify the influence of single and multi-purpose functions on community detection performance. Finally, this paper gives a comparative analysis of proposed MPPOA algorithm worth three heuristic overlap community detection algorithm over six real social media data set.

Keywords: Classification, Community detection, Eligibility values, Multi-purpose function, Parliamentary optimization algorithm, Seed community, Single function, Social media.

Introduction

The purpose of the World Wide Web (WWW) is to provide information to users through websites. In Web 1.0 sites, this presentation is one-way, and users are passive. On Web 2.0 sites, some active users produce and share content. Web 2.0 tools that use social interaction possibilities are called social media. In these environments, users use social networking sites, blogs (blogs), wikis (knowledge pages) and forums (discussion boards) to share and discuss their experiences, knowledge. [1, 2] Among the general features of social media; participation, openness, conversation, community, and connectivity.[3,4] Social networking sites are web-based services that enable individuals to create profiles and connect with other userswithin a certain system. Profile presentation indicates how they connect with other users (such as friends, fans, followers); It provides both communications with new people and meeting of acquaintances. The formation of virtual communities in these sites means that mobility and interaction are based on user performances, which increase information dissemination. In social networks, the efforts of individuals to As a result of the research on the overlapping community[8–10] discovery problem in social networks, it has been seen that many of the algorithms developed

previously for community discovery solve this problem by using a single goal. At the same time, it has been determined that there are many newly discovered and proposed socialbased algorithms.[11–14] One of these algorithms is the Parliamentary Optimization Algorithm (POA).[13] POA simulates real-life parliamentary elections.[15] The optimization process in the algorithm begins with the creation of the individual population first. These individuals are considered members of Parliament. In the next step, the population is distributed among some political groups, and a fixed number of highly suiTable members are selected as candidates for the group. The principal members of the group head towards the candidate members, and after the orientation process, the candidate and leading members of the group are recalculated. The calculated new candidates and principal members are used to calculating the strengths of the groups. Strong groups join forces, while vulnerable

groups are wiped out to keep their power from waning. When the termination condition of the algorithm is met, the best individual solution to the optimization problem in the population is accepted. The freshly proposed POA was used within the scope of this research paper. The specified algorithm has not been used previously for the problem of overlapping community discovery in social networks. The algorithm has been applied for the first time in this work to discover communities that overlap in social networks; that is, a community member can be included in another district by using both single-purpose and multi-purpose functions.[14] With the new method developed, the modularity of social networks was provided by using a single-purpose part. The proposed method has been tested on a synthetic dataset. Then, a single-purpose algorithm developed by adding a new objective function to optimize the internal density in communities in the network was transformed into a multi-purpose form. Thus, using POA and multipurpose optimization together, the first proposed algorithm for overlapping community discovery problems in social networks was developed. The rest of the paper is organized as follows: Section.2

gives bird's eye over the Parliamentary optimization Algorithm; Section.3 presents Single and Multi-purpose function-based community detection Methods; Section.4 covers the experimental set-up and performance evaluation of the Proposed overlapping community detection algorithm and finally, Sect.[8] concludes the paper and outlines the founding and future work. Parliamentary Optimization Algorithm (POA) The parliamentary system, a system of government making and regulating laws, is also known as parliament tsarism. The people elect members of Parliament in general elections. People often vote for their favorite party. Members of Parliament who are members of political parties support their parties in parliamentary elections. Parliamentary groups of
members based on the party they belong to strive to gain superiority over other parties in the competition between parties. In almost all democratic countries, the parliamentary population is formed by political parties. There are two systems in parliamentary elections, the majority election system and the proportional representation system. While only one member is elected from each constituency in the majority electoral system, several members may be selected from one constituency in the balanced representation system. Generally, each political party presents their list of candidates, and voters can choose the political parties to vote for. Parties are given seats in the Parliament in proportion to their votes.[16] Members of political parties within or outside Parliament have different power values. These members of the party strive with little power to make a good impression on other noble members. They make this effort to get their support and votes during the elections. Essential members of the party get involved in races and try to find support among the
noble members. On the other hand, Noble members tend to be more resourceful and often vote for those they trust. Inthis process, high-capacity general members are replaced with previous candidates. This part of the competition takes place between individuals within the party. Another race of the algorithm takes place between parties. Parties complete to increase their power. Parties have two main objectives for success: having the highest number of seats in Parliament and taking control of the government.[16] In the Parliamentary Optimization Algorithm (POA), optimization steps begin with creating the initial population of individuals. The individuals created are considered members of the Parliament. In the next step, the population is divided into political groups (parties), and the candidate for the fixed number of member groups with the highest fitness is considered. After this step, the in-group competition starts. In the in-group competition step, the leading members turn to the candidate members suiTable for them. This situation is modelled as the weighted average of vectors of principal
candidates,[16] as shown inthe POA flow chart, i.e., Figure 1. At the end of the in-group competition step, several

Journal of Namibian Studies, 33 S1 (2023): 3826-3852 ISSN: 2197-5523 (online)

candidates with the highest qualification are determined as the final candidates for each group. In the next step,
the final candidates compete with the candidates of other groups. Principal and candidate members of the group are
essential in determining the total power of the group. After the intra-group competition step, the competition between groups starts. Political groups within Parliament compete with other groups to strengthen their candidates. Strong groups sometimes unite and become one group to increase their chances of winning. Algorithm 1 shows the process steps of POA

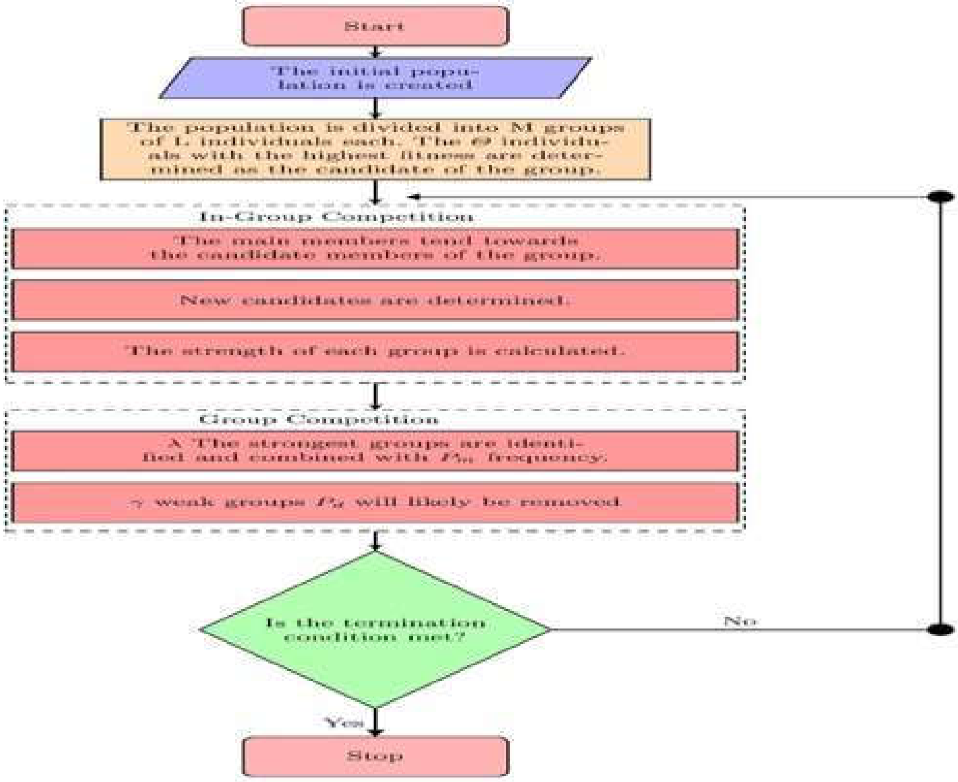
Algorithm 1 Stepwise explanation of Parliamentary Optimization Algorithm(POA):

• Start

• The initial population is created.

(a)The population is divided into M groups consisting of L individuals.

(b)The highly fit individual is selected as the candidate for each group.



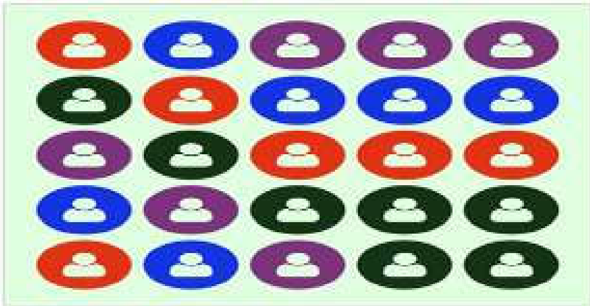


Figure 2: Segmentation of the population

(c)In-group competition

(a)The prominent members head towards the candidate members of each group.

(b)New candidates are appointed.

(c)Calculate the power of each group.

• Competition between groups

(a)The most influential group is determined, and these groups are combined with Pm probability.

(b)The weakest group Pd will likely be deleted.

• If the termination condition is not met, step 3 is repeated.

• The best candidate is considered the solution to the optimisation problem.

• Stop.

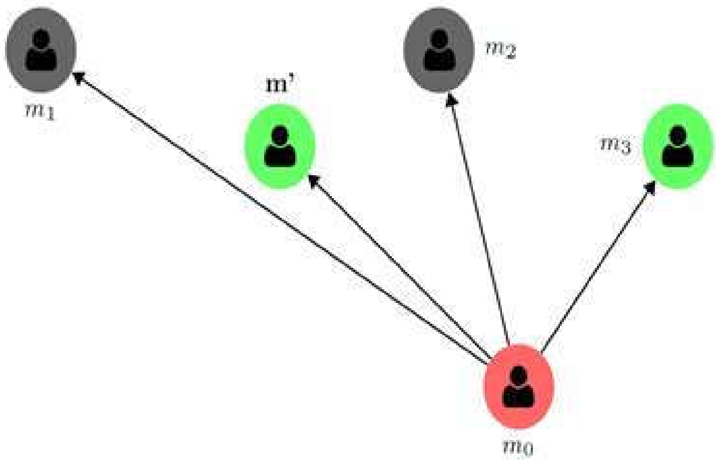


Figure 3: Orientation Mechanism determined

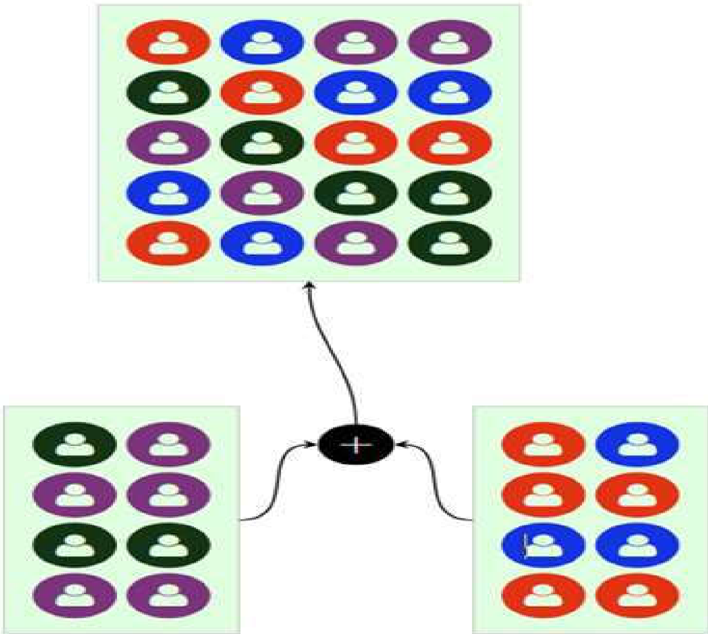


Figure 4: Joining Of Group

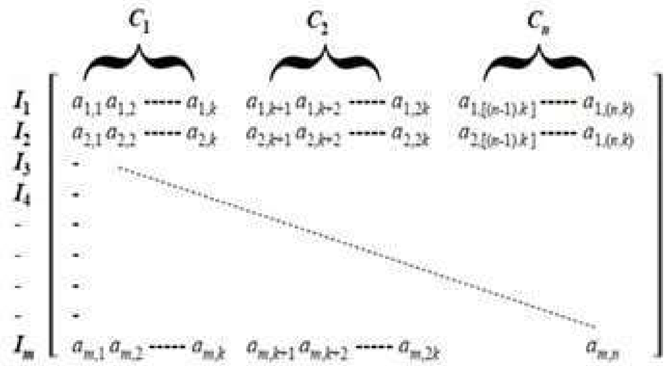


Figure 5: Representation of the initial population

Journal of Namibian Studies, 33 S1 (2023): 3826-3852 ISSN: 2197-5523 (online)

Each individual is the principal member or candidate of the given group. Individuals' strengths are calculated according to the determined fitness function.

Population Segmentation

To form the starting groups, the population is divided into M groups of L individuals.

$N = M * L$ (2)

N is as in Equation 2, where N, M and L are positive integers. High fidelity $\theta < L/3$ candidates are determined as candidates for the groups. In this step of the algorithm, all groups have an equal number of members. During the algorithm's operation, groups can obtain different numbers of individuals due to the merger and collapse mechanism. Figure 2 shows the division of an initial population into 3 groups of 5 candidates each. The solid blue symbols in the Figure 2 represent the candidates of the group.

In-Group Competition

After replacing the candidate and the leading members, the group's principal members head towards the candidates. The orientation process is directly proportional to the weighted averages of the vectors connecting a member to candidates. Each candidate is weighted to increase his or her candidate eligibility, as shown in Equation 3.

$m' = m_0 + \eta \frac{\Sigma^{*}_{i}(m_i - m_0).f(m_i)}{\Sigma^{*}_{i} f(m_i)}$ (3)

In the formula, η is a random value ranging from 0.5 to 1, allowing the algorithm to search for candidates within the local search space. A principal member is replaced only if the fitness value is high. After the referral process, the fitness value of top members is higher than that of candidate members. Figure 3 shows the orientation mechanism. m_0 is a full member, and m_i is a candidate member. m is the new position of the leading member.

The vector of eligibility value of candidates belong to community $ev_i = \{ec_{i,1}, ev_{i,2}, \ldots, ev_{i,\theta}\}$ and eligibility value of candidates; $ev_{ci} = ev_{ci,\theta+1}, ev_{ci,\theta+2}, \ldots, ev_{ci,Li}$.

The strength of the group, including the principal members of the group, is calculated by Equation 4.

$Strength^{i} = \frac{m.Avg(ev_i) + n.Avg(ev_i^c)}{m+n}; m \geq n$ (4)

In the equation, m and n are the weighted constants of the candidate and principal members.

2.4 Competition between Groups Strong groups sometimes join and unite in a group to increase their strength. A random number is generated to accomplish the merger, and if this number is less than pm, the most influential group in λ number is determined and combined into a group. The vulnerable groups are deleted during the algorithm to maintain the power value and reduce the value function. Figure 4 shows the merger of the two groups. As in combining, a random number is generated, and if the number is less than PD, groups with a minimum power of y are eliminated.[17]

Termination of the situation

At the end of the algorithm, a group wins the race, and the best member of this group is considered the solution to the optimisation problem. There are two cases of termination: The algorithm is terminated when the maximum number of iterations is reached or if no significant improvement in the fitness value is observed due to some successful iteration.[17]

Multi-purpose Overlapping Community Detection With POA (MPPOA)

Objective functions play an important role in optimisation problems so that multiple purposes will be used with POA. Many objective functions have been proposed, especially for community discovery. If G (V, E) is accepted as a non-directional chart; $n = |V|$ and $m = |E|$ happens. S is the set of nodes in the group, and k is the number of nodes in the set $S. k = |S|$ is the number of sides in the set $S. l = |\{(u,v): u \varepsilon S, v \varepsilon S\}|$

cS is the number of edges within the boundaries of the set S. $cS = |\{(u,v): u \varepsilon S, v \notin S\}|$ $d(u)$ is the degree of node d. Some objective functions that measure the quality concept of a cluster using the given definitions are given below18

Conductance : A metric that calculates the total volume of links pointing out of the cluster.

$f(S) = \frac{c_S}{(2.1 + c_S)}$ (5)

Expansion : Returns the average number of external links for each node in the cluster.

$f(S) = \frac{c_S}{k}$ (6)

Internal Density : S is the density of the internal links in the cluster.

$f(S) = \frac{1}{(k(k-1)/2)} -$ (7)

Cut Radio : It is the ratio of all possible links leaving the cluster.

$f(S) = \frac{c_S}{k(n-k)}$ (8)

Normalized Cut :

$f(S) = \frac{c_S}{(2.1 + c_S)} + f(S) = \frac{c_S}{2.(m-1) + c_S}$ (9)

Maximum ODF : It is the ratio of external links to internal links for each node in cluster S.

$\text{maximum } u \varepsilon S \frac{|\{(u,v): v \notin S\}|}{d(u)}$ (10)

Average ODF : It is the average ratio of connections of nodes outside of the cluster.

$f(S) = \frac{1}{k} \Sigma u \varepsilon S \frac{|\{(u,v): u \notin S\}|}{d(u)}$ (11)

Flake ODF : It is the ratio of nodes in the S cluster with fewer connections outside the cluster than inside the cluster.

$f(S) = \frac{|\{u: u \varepsilon S, |\{(u,v): v \varepsilon S\}| < d(u)/2\}|}{k}$ (12)

As mentioned in the previous section, POA is initiated with the creation of the initial population. This population of individuals is considered to be members of Parliament. Individuals produced between 0 and 1 $(I_1, I_2, \ldots, I_m)$ form the initial population. The starting population is expressed as given in Figure 5.

In the matrix, k is the total number of nodes in the network. In this step, the population is divided into M groups

Average ODF : It is the average ratio of connections of nodes outside of the cluster.

(11) Flake ODF : It is the ratio of nodes in the S cluster with fewer connections outside the cluster than inside the cluster.

(12) As mentioned in the previous section, POA is initiated with the creation of the initial population. This population of individuals is considered to be members of Parliament. Individuals produced between 0 and 1 form the initial population. The starting population is expressed as given in Figure 5. In the matrix, k is the total number of nodes in the network. In this step, the population is divided into M groups

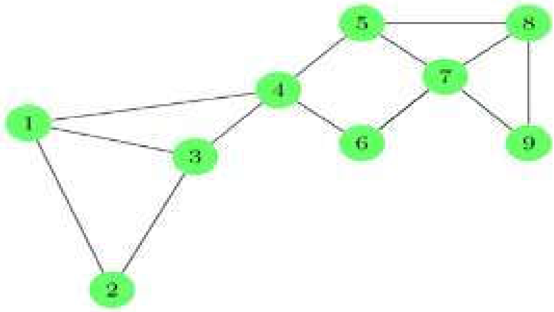


Figure 6: A typical network structure of L individuals each. At this stage, a multi-purpose method was used to determine the candidates of the group. One of the objective functions is modularity in the network. Expanded modularity has been proposed to identify overlapping communities.11 Modularity allows measuring the quality of specific sections in the network. Calculates the strength of the network by comparing the distribution of links within the network. Extended modularity is given in Equation 13. ……………(13)

Cni given in the equation is the number of ensembles to which the node v belongs, is the element of the neighbour matrix of the network. value is 1 if there is a connection between i and j nodes; otherwise, it takes deal 0. di P is the degree of node i, and e is the total number of links in the network. It has been observed in the literature that a high modularity value indicates strong community structures. For this reason, it is expected that the value will take its maximum value in this multi-purpose study. Another objective function used in this step of the algorithm is the internal density criterion given in Equation 7. A small value must be obtained from the given equation to get a strong ensemble structure. (14) Where is the total number of node in network? In this step of the POA, a multi-purpose method has been proposed to determine the group's candidates by combining two objective functions (Equation 15). The value of a is an input value used to emphasize one of the purposes. Those with high-cost value in the equation are accepted as candidates for the groups. In the in-group competition step, the group's permanent members turn to the candidate members according to Equation 3. After the orientation process, the candidate and principal members of the group are determined again, and the power of the group is calculated according to

Equation 4. After the intra-group competition step, powerful groups unite in one group to
increase their strength. If a significant increase in algorithm steps is not observed, the termination condition is met. At the end of the algorithm, a group wins the race, and the best member of that group is considered the solution to the multipurpose overlapping community discovery. (15) Comprehensive Analysis of Proposed Algorithm In this paper, a multi-purpose method for community discovery with POA is proposed, and the experimental results for the proposed algorithm are tested on different data sets. POA was carried out for a single purpose.[17] In this part of the study, the program implemented in Python environment for both a single goal and a multi-purpose approach was applied to artificial and real network data. Single-Purpose Community Discovery with POA on Artificial Dataset For the proposed single-purpose study, first of all, the way the data is represented was determined. In Figure 5, the data representation format used for the multi-purpose method is also used for the single-purpose approach. The effectiveness of the study was first tested on the designed artificial network. The experiment was carried out to measure the ability of POA on community discovery. An artificial network consisting of 9 nodes and 13 connections is given in Figure 6. In the first step of the algorithm, the starting population was created. These values were generated in the Python environment. The first population produced is divided into 3 groups of 10 individuals each. In this case, the importance of the variables in Equation 2 is given in Table 1.
The values of the parameters in Equation 3 and 4 in the intra-group competition step of the POA are given in Table 2. In the single-purpose approach, the objective function used to find the fitness values of the individuals in the in-group competition step is the formula given in Equation 13. This formula, called extended modularity, calculates the strength of the network by calculating the distributions of the links in the network. The highest 3 individuals among the calculated eligibility values are accepted as candidates of the group. The fitness values of the individuals of each group are given in Table 3. Values written in bold in the Table are accepted as candidates for the groups. According to Equation 3, other members of the group head towards the candidate of the group and new candidates are determined. After determining the new candidates, the of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table, the 1st individual is considered the solution to the problem because he has the highest fitness. The communities found by the POA for the artificial data set are given in Figure 7. The algorithm seen in Figure 7 has found two groups. Nodes 5 and 6 belong to both clusters and overlap. The communities to which the nodes belong are given in Table 6. Nodes in bold represent overlapping knots. Multi-purpose Community Overlap Discovery with POA on

Artificial Dataset The efficiency of the proposed multi-purpose study was first tested on the designed artificial network. The experiment

Variables	*Values*
N	30
M	3
L	10

Table 2: Values of parameters in the intra-group competition step

Parameters	*Values*
η .	68
m .	58
n .	23

Table 3: Eligibility values of Individuals in Groups of the Artificial Data Set.

Group 1		*Group 2*		*Group 3*	
I@	*EV#*	*I*	*EV*	*I*	*EV*
I1	2	I11	1.5	I21	1.1
I2	1.5	I12	0.8	I22	0.2
I3	1.4	I13	0.9	I23	0.7
I4	2	I14	0.7	I24	1
I5	2.7	I15	2.1	I25	1.4
I6	2.9	I16	0.9	I26	2.4
I7	2.3	I17	0.3	I27	0
I8	1.5	I18	0.4	I28	1.1
I9	0.6	I19	2.1	I29	1.3
I10	0.6	I20	2	I30	1.3

Table 4: Strengths of the Groups in the synthetic dataset

Groups	*Strengths of the Groups*
1	2.26
2	1.57
3	1.41

Table 5: Individual Eligibility values in the Artificial Data Set

I@	EV#	I	EV	I	EV
I1	4.6	I11	1.5	I21	1.4
I2	3.8	I12	2.1	I22	1.3
I3	3.8	I13	0.9	I23	1.1
I4	3.8	I14	2.1	I24	1.4
I5	2.7	I15	3.4	I25	1.4
I6	4.6	I16	2.1	I26	2.4
I7	3.8	I17	2.1	I27	1.1
I8	3.8	I18	1	I28	1.4
I9	3.8	I19	2.1	I29	1.4
I10	3.8	I20	3.4	I30	1.1

Table 6: The communities found for the synthetic dataset.

Communities	Nodes
1	1,2,3,4,5,6
2	5,6,7,8,9

Table 5: Individual Eligibility values in the Artificial Data was carried out to measure the ability of POA on community discovery. An artificial network consisting of 9 nodes and 13 connections is given in Figure 8. In the first step of the algorithm, the starting population was created. These values were generated in the Python environment. The first population produced is divided into 3 groups of 10 individuals each as shown in Figure 10. According to the formula given in Equation 15, the fitness values of each individual in the group were calculated. The highest 3 individuals among the calculated eligibility values are accepted as candidates of the group. The fitness values of the individuals of each group. Values written in bold in the Table are accepted as candidates for the groups. According to Equation 3, other members of the group head towards the candidate of the group and new candidates are determined. After determining the new candidates, the Strengths of each group was calculated, and the values. After the intra-group competition step, the competition between groups starts. In this step, the strongest λ = 2 groups are combined at Pm = 30% or deleted for Pd = 1%. If the groups do not merge, return to the intra-group competition step. These steps continue until all groups are united or the

best solution to the problem has been achieved. After all, groups are combined, the eligibility values of the individuals are given in Table 9. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table , the 2nd individual is considered the solution to the problem because he has the highest fitness. The communities found by the POA for the artificial data set are given in Figure 9. The algorithm shown in Figure 9 has found two groups. Nodes 5 and 6 belong to both clusters and overlap. The communities to which the nodes belong. Nodes in bold represent overlapping knots. Multipurpose Community Overlap Discovery with POA on Real-World Data

In this part of the study, the proposed algorithm was tested on 4 different datasets, including Zachary's Karate Club,[19] American College Football,[20] Dolphin Social Network,[21] Lesmis[22] and to discover overlapping communities were used as a multi-purpose approach. Zachary's Karate Club Data Set Zachary's Karate Club is a social network that shows the friendship relationships between 34 members of the karate club at US University in 1970. For this network consisting of 78 connections, the initial population, the first step of POA, was created in the Python environment. The first population produced is divided into 2 groups The fitness value of each individual in the group is calculated according to the cost value in Equation 15, and the principal and candidate members of the group are determined. Principal and candidate members are given in Table 7. Individuals written in bold are candidate members of the groups. After the candidate members are determined, the competition step within the group begins, and the permanent members of the group head towards the candidate members. After the orientation process, the candidate and principal members of the group are recalculated. The strengths of the groups are calculated using Equation 4 according to the determined candidates and leading members. Power values are given in Table 8. After the intra-group competition step, the competition between groups starts. In this step, the strongest λ = 2 groups are Table 7: Eligibility values of individuals in groups in Zachary's Karate Club dataset Eligibility Values

Table 7: Eligibility values of individuals in groups in Zachary's Karate Club dataset

Eligibility Values	*Individuals in Community*									
	I1	*I2*	*I3*	*I4*	*I5*	*I6*	*I7*	*I8*	*I9*	*I10*
Community1	67.392	85.086	71.01	78.48	81.468	105.57	80.289	89.154	86.859	58.077
Community2	103.275	89.352	89.721	10.341	60.876	54.81	58.356	82.152	39.474	59.346

Table 8: Strengths of the Groups in Zachary's Karate Club Data Set

Groups	*Strengths of the Groups*
1.	98.07
2.	91.07

Multipurpose Community Overlap Discovery with POA on Real-World Data

In this part of the study, the proposed algorithm was tested on 4 different datasets, including Zachary's Karate Club,[19] American College Football,[20] Dolphin Social Network,[21] Lesmis[22] and to discover overlapping communities were used as a multi-purpose approach.

Zachary's Karate Club Data Set

Zachary's Karate Club is a social network that shows the friendship relationships between 34 members of the karate club at US University in 1970. For this network consisting of 78 connections, the initial population, the first step of POA, was created in the Python environment. The first population produced is divided into 2 groups The fitness value of each individual in the group is calculated according to the cost value in Equation 15, and the principal and candidate members of the group are determined. Principal and candidate members are given in Table 7. Individuals written in bold are candidate members of the groups. After the candidate members are determined, the competition step within the group begins, and the permanent members of the group head towards the candidate members. After the orientation process, the candidate and principal members of the group are recalculated. The strengths of the groups are calculated using Equation 4 according to the determined candidates and leading members. Power values are given in Table 8. After the intra-group competition step, the competition between groups starts. In this step, the strongest $\lambda = 2$ groups are

Table 9: Eligibility values of individuals in Zachary's Karate Club dataset

Eligibility Values	*Individuals in Community*									
	I1	*I2*	*I3*	*I4*	*I5*	*I6*	*I7*	*I8*	*I9*	*I10*
Community1	98.478	108.49	70.254	121.96	35.505	84.865	48.321	106.73	107.76	114.88
Community2	103.275	89.352	89.721	10.341	60.876	54.81	58.356	82.152	39.474	59.346

Table 10: Communities found for Zachary's Karate Club dataset

Communities	*Nodes*
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,18, 20, 21, 22, 29, 30, 31
2	9, 15, 16, 17, 19, 20, 23, 24, 25, 26, 27,28, 29, 30, 31, 32, 33, 34

Table 11: Eligibility values of individuals in groups in the American College Football dataset

Eligibility Values	*Individuals*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	112.36	116.35	101.2	98.81	93.23	87.65	82.07	76.49	70.91	65.33
Community-2	56.96	59.36	55.63	55.99	55.32	54.66	53.99	53.33	52.66	52
Community-3	89.45	96.45	94.36	98.33	100.79	103.24	105.7	108.15	110.61	113.06
Community-4	124.56	135.36	123.36	126.56	125.96	125.36	124.76	124.16	123.56	122.96
Community-5	56.78	69.23	68.32	76.32	82.09	87.86	93.63	99.4	105.17	110.94
Community-6	89.66	96.45	92.36	95.52	96.87	98.22	99.57	100.92	102.27	103.62
Community-7	79.63	89.32	82.36	86.5	87.87	89.23	90.6	91.96	93.33	94.69
Community-8	96.85	97.35	95.36	95.03	94.29	93.54	92.8	92.05	91.31	90.56
Community-9	116.23	125.36	121.36	126.11	128.68	131.24	133.81	136.37	138.94	141.5
Community-10	165.32	125.96	153.63	136.61	130.77	124.92	119.08	113.23	107.39	101.54

solution to the problem has been achieved. After all groups are combined, the fitness values of individuals. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table , the 9th individual is considered

the solution to the problem because he has the highest fitness. The communities found by POA for Zachary's Karate Club are given in Figure 11.

MPPOA Framework for Community Detection over Social Media

Table 12: Strengths of the groups in american college football data set

Groups	Strengths of the Groups
1.	18.90
2.	14.29
3.	17.32
4.	12.10
5.	14.12
6.	26.02
7.	16.90
8.	11.18
9.	13.67
10.	14.78

combined at Pm = 30% or deleted for Pd = 1%. If the groups do not merge, return to the intra-group competition step. These steps continue until all groups are united or the best solution to the problem has been achieved. After all groups are combined, the fitness values of individuals. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem.

According to the Table , the 9th individual is considered the solution to the problem because he has the highest fitness. The communities found by POA for Zachary's Karate Club are given in Figure 11.

The algorithm suggested for Zachary's Karate Club found 2 communities. While the green and purple coloured nodes represent these 2 groups, the blue coloured nodes 9, 20, 29, 30 and 31 indicate the overlapping nodes belonging to both communities. The communities to which the nodes belong are given in Table 10. Overlapping nodes are written in bold in the Table .

American College Football Dataset

Nodes in the American College Football network link football teams, while the games played between the two groups during the 2000 football season. For this network consisting

Table 13: Eligibility values of individuals in the American College Football data set

Eligibility Values	*Individuals*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	98.88	102.39	89.06	86.95	82.04	77.13	72.22	67.31	62.4	57.49
Community-2	50.12	52.24	48.95	49.27	48.68	48.1	47.51	46.93	46.34	45.76
Community-3	78.72	84.88	83.04	86.53	88.69	90.85	93.01	95.17	97.33	99.49
Community-4	109.61	119.12	108.56	111.37	110.84	110.32	109.79	109.26	108.73	108.2
Community-5	49.97	60.92	60.12	67.16	72.24	77.31	82.39	87.47	92.55	97.62
Community-6	78.9	84.88	81.28	84.06	85.25	86.44	87.62	88.81	90	91.19
Community-7	70.07	78.6	72.48	76.12	77.32	78.52	79.72	80.92	82.13	83.33
Community-8	85.23	85.67	83.92	83.63	82.97	82.32	81.66	81	80.35	79.69
Community-9	102.28	110.32	106.8	110.98	113.24	115.49	117.75	120.01	122.27	124.52
Community-10	145.48	110.84	135.19	120.22	115.08	109.93	104.79	99.65	94.5	89.36

Table 14: Eligibility values of the individuals in the groups in the Dolphin Social Network dataset

Eligibility Values	*Individuals in Community*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	129.21	133.8	116.38	113.63	107.21	100.8	94.38	87.96	81.55	75.13
Community-2	65.5	68.26	63.97	64.38	63.62	62.86	62.09	61.33	60.56	59.8
Community-3	102.87	110.92	108.51	113.08	115.9	118.73	121.55	124.37	127.2	130.02
Community-4	143.24	155.66	141.86	145.54	144.85	144.16	143.47	142.78	142.09	141.4
Community-5	65.3	79.61	78.57	87.76	94.4	101.04	107.67	114.31	120.94	127.58
Community-6	103.11	110.92	106.21	109.85	111.4	112.96	114.51	116.06	117.61	119.17
Community-7	91.57	102.72	94.71	99.48	101.04	102.61	104.18	105.75	107.32	108.89
Community-8	111.38	111.95	109.66	109.28	108.43	107.57	106.71	105.86	105	104.14
Community-9	133.66	144.16	139.56	145.03	147.98	150.93	153.88	156.83	159.78	162.73
Community-10	190.12	144.85	176.67	157.11	150.38	143.66	136.94	130.22	123.5	116.77

The algorithm suggested for Zachary's Karate Club found 2 communities. While the green and purple coloured nodes represent these 2 groups, the blue coloured nodes 9, 20, 29, 30 and 31 indicate the overlapping nodes belonging to both communities. The communities to which the nodes belong are given in Table 10. Overlapping nodes are written in bold in the Table . American College Football Dataset Nodes in the American College Football network link football teams, while the games played between the two groups during the 2000 football season. For this network consisting

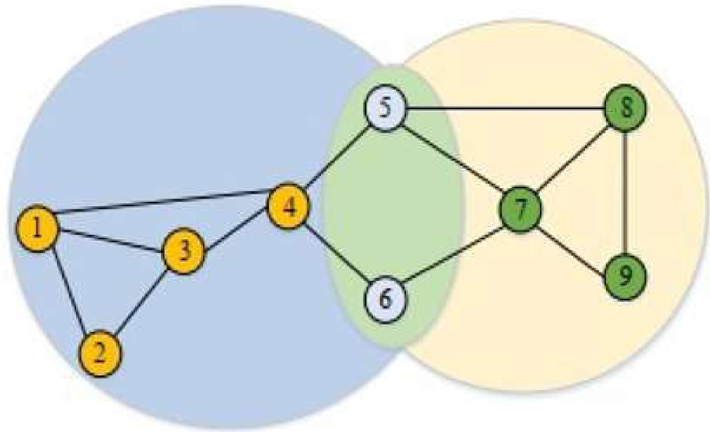


Figure 7: Communities found by POA for synthetic dataset

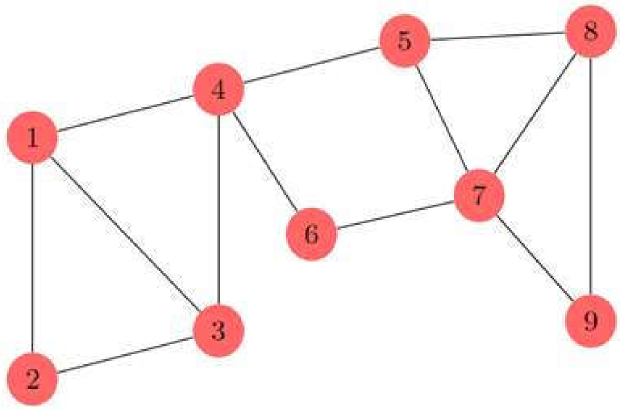


Figure 8: Artificial Network created in Pajek Environment

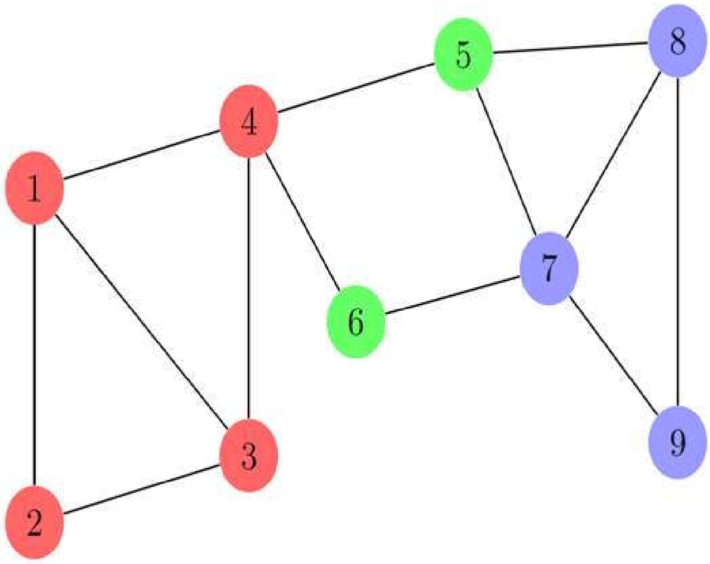


Figure 9: Network Structure of 115 nodes and 610 connections, the initial population, which is the first step of POA, was created in the Python environment. The first population produced is divided into 10 groups . The eligibility values of each individual in the group is calculated according to the cost value in Equation 5.10, and the principal and candidate members of the group are determined. Principal and candidate members are given in Table 11. Individuals written in bold are candidate members of the groups.

p

MPPOA Framework for Community Detection over Social Media

Table 16: Eligibility values of individuals in the Dolphin Social Network dataset

Eligibility Values	*Individuals*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	151.18	156.55	136.16	132.95	125.44	117.93	110.43	102.92	95.41	87.9
Community-2	76.64	79.87	74.85	75.33	74.44	73.54	72.65	71.75	70.86	69.96
Community-3	120.35	129.77	126.96	132.3	135.61	138.91	142.21	145.52	148.82	152.12
Community-4	167.6	182.13	165.98	170.29	169.48	168.67	167.86	167.06	166.25	165.44
Community-5	76.4	93.15	91.92	102.68	110.45	118.21	125.97	133.74	141.5	149.27
Community-6	120.64	129.77	124.27	128.53	130.34	132.16	133.98	135.79	137.61	139.43
Community-7	107.14	120.18	110.82	116.39	118.22	120.06	121.9	123.73	125.57	127.41
Community-8	130.31	130.98	128.31	127.86	126.86	125.86	124.86	123.85	122.85	121.85
Community-9	156.39	168.67	163.29	169.69	173.14	176.59	180.04	183.49	186.94	190.39
Community-10	222.44	169.48	206.71	183.81	175.95	168.08	160.22	152.36	144.49	136.63

Table 17: Eligibility values of the individuals in the groups in the Lesmis data set

Eligibility Values	*Individuals in Community*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	232.59	240.84	209.48	204.54	192.99	181.44	169.88	158.33	146.78	135.23
Community-2	117.91	122.88	115.15	115.89	114.52	113.14	111.76	110.39	109.01	107.63
Community-3	185.16	199.65	195.33	203.54	208.62	213.71	218.79	223.87	228.95	234.03
Community-4	257.84	280.2	255.36	261.98	260.74	259.5	258.25	257.01	255.77	254.53
Community-5	117.53	143.31	141.42	157.98	169.92	181.86	193.81	205.75	217.7	229.64
Community-6	185.6	199.65	191.19	197.73	200.53	203.32	206.12	208.91	211.71	214.5
Community-7	164.83	184.89	170.49	179.06	181.88	184.71	187.53	190.36	193.18	196.01
Community-8	200.48	201.51	197.4	196.71	195.17	193.63	192.09	190.54	189	187.46
Community-9	240.6	259.5	251.22	261.05	266.36	271.67	276.98	282.29	287.6	292.91
Community-10	342.21	260.74	318.01	282.79	270.69	258.59	246.49	234.39	222.29	210.19

Table 18: Eligibility values of individuals in the Lesmis data set

Eligibility Values	*Individuals*									
	I-1	*I-2*	*I-3*	*I-4*	*I-5*	*I-6*	*I-7*	*I-8*	*I-9*	*I-10*
Community-1	204.67	211.94	184.35	179.99	169.83	159.66	149.5	139.33	129.17	119.01
Community-2	103.76	108.13	101.34	101.99	100.77	99.56	98.35	97.14	95.93	94.72
Community-3	162.94	175.69	171.89	179.12	183.59	188.06	192.53	197.01	201.48	205.95
Community-4	226.9	246.57	224.71	230.54	229.45	228.36	227.26	226.17	225.08	223.98
Community-5	103.43	126.11	124.45	139.02	149.53	160.04	170.55	181.06	191.57	202.08
Community-6	163.32	175.69	168.24	174.01	176.46	178.92	181.38	183.84	186.3	188.76
Community-7	145.05	162.71	150.03	157.57	160.05	162.54	165.03	167.51	170	172.49
Community-8	176.42	177.33	173.71	173.11	171.75	170.39	169.04	167.68	166.32	164.96
Community-9	211.72	228.36	221.07	229.73	234.4	239.07	243.75	248.42	253.09	257.76
Community-10	301.15	229.45	279.85	248.85	238.21	227.56	216.91	206.27	195.62	184.97

After the candidate members are determined, the competition step within the group begins, and the permanent members of the group head towards the candidate members.

MPPOA Framework for Community Detection over Social Media

Table 19: Comparative Analysis of Impact of Social theory on modularity

Classification Technique	*Modularity*					
	ZKC	*ACF*	*DCN*	*BUP*	*LM*	*WA*
SEOA	0.5313	0.6107	0.661	0.6151	0.6121	0.4118
ICA	0.4302	0.7211	0.6179	0.6714	0.6122	0.4192
TLBO	0.5137	0.6204	0.6196	0.7119	0.6217	0.5103
MPPOA	0.7456	0.8056	0.7435	0.8123	0.7658	0.6856

Table 20: Comparative Analysis of Impact of Social theory on Normalized Mutual Information

Classification Technique	*Normalized Mutual Information*					
	ZKC	*ACF*	*DCN*	*BUP*	*LM*	*WA*
SEOA	0.8102	0.8233	0.7162	0.6955	0.5253	0.5822
ICA	0.8651	0.6324	0.5861	0.5712	0.4128	0.4268
TLBO	0.8624	0.7823	0.7152	0.5251	0.4462	0.5122
MPPOA	0.9087	0.8763	0.8567	0.7652	0.7459	0.7125

After the orientation process, the candidate and principal members of the group are recalculated. The strengths of
the groups are calculated using Equation 4 according to the determined candidates and leading members. Strengths of the Groups are given in Table 12. After the intra-group competition step, the competition between groups starts. In this step, the strongest $\lambda = 2$ groups are combined at Pm = 30% or deleted for Pd = 1%. If the groups do not merge, return to the intra-group competition step. These steps continue until all groups are united or the best solution to the problem has been achieved. After all, groups are combined, the eligibility values of the individuals are given in Table 13. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table , the 2nd individual is considered the solution to the problem because he has the highest fitness. The communities found by POA for American College Football are given in Figure 12. The algorithm suggested for American College Football has found 3 communities. Green, Yellow and Red coloured nodes represent these 3 groups, while the blue-, pink- and purplecoloured knots show the overlapping nodes belonging to 2 or 3 communities. Dolphin Social Network Data set This dataset is Doubtful Sound, a non-directional network of frequent relationships among 62 dolphins living in a closed New Zealand community. For this network consisting of 159 connections, the initial population, which is the first step of POA, was created in the Python environment. The first population produced is divided into 3 groups of 10 individuals each. The eligibility values of each individual in the group is calculated according to the cost value in Equation 15, and the principal and candidate members of the group are determined. Principal and candidate members are given in Table 14. Individuals written in bold are candidate members of the groups. After the candidate members are determined, the competition

step within the group begins, and the permanent members of the group head towards the candidate members. After the orientation process, the candidate and principal members of the group are recalculated. The strengths of the groups are calculated using Equation 4.4 according to the determined candidates and leading members. Strengths of the Groups are given in Table 15. After the intra-group competition step, the competition between groups starts. In this step, the strongest λ =2 groups are combined at Pm = 30% or deleted for P d= 1%. If the groups do not merge, return to the intragroup competition step. These steps continue until all groups are united or the best solution to the problem has been achieved. After all, groups are combined, the eligibility values of the individuals are given in Table 16. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table , the 2nd individual is considered the solution to the problem because he has the highest fitness. Communities found by POA for Dolphin

Social Network are given in Figure 13. The algorithm suggested for Dolphin Social Network has found 2 communities. The green- and purple-coloured nodes represent these 2 groups, while the blue coloured nodes indicate the overlapping nodes belonging to both communities 4.3.4 Lesmis Dataset Lesmis shows the collaboration of 77 characters in Victor Hugo's novel Les Miserables. For this network consisting of 254 connections, the initial population, which is the first step of POA, was created in the python environment. The first population produced is divided into 5 groups of 10 individuals each. The eligibility values of each individual in the group is calculated according to the cost value in Equation 15 and the principal and candidate members of the group are determined. Principal and candidate members are given in Table 17. Individuals written in bold are candidate members of the groups. After the candidate members are determined, the competition step within the group begins, and the permanent members of the group head towards the candidate members After the orientation process, the candidate and principal members of the group are recalculated. The strengths of the groups are calculated using Equation 4 according to the determined candidates and leading members. After the intra-group competition step, the competition between groups starts. In this step, the strongest λ =2 groups are combined at Pm = 30% or deleted for Pd = 1%. If the groups do not merge, return to the intragroup competition

step. These steps continue until all groups are united or the best solution to the problem has been achieved. After all, groups are combined, the eligibility values of the individuals are given in Table 18. As stated in the termination condition of the POA, the highest value in the Table is accepted as the solution to the multi-purpose overlapping community discovery problem. According to the Table , the 9th individual is considered the solution to the problem because he has the highest fitness. The communities found by POA for Lesmis are given in Figure 14. The algorithm suggested for Lesmis has found 2 communities. The green and purple coloured nodes represent these 2 groups, while the blue coloured nodes indicate the overlapping nodes belonging to both communities. Experimental Set up and Result Analysis Performance evaluation to detect the impact of single and multi-purpose based heuristic community detection algorithm has been carried out over six different graphical

social media data sets, namely Word adjacencies, Zachary karate club,[19] Dolphin social network,[20] Les Miserables, Books about US politics and American College football[21] over the evaluation parameter modularity and normalized mutual information. Modularity is network structural measurement that evaluate the strength of sub graph (groups, clusters or communities) in network for extracting community structure. [23] In a network, group of node having higher modularity are

	Community 1									Community 2								
I_1	0.28	0.49	0.34	0.78	0.36	0.73	0.38	0.78	0.67	0.48	0.74	0.64	0.75	0.45	0.56	0.46	0.60	0.34
I_2	0.43	0.74	0.03	0.94	0.76	0.55	0.18	0.49	0.51	0.99	0.85	0.96	0.67	0.40	0.93	0.47	0.23	0.39
I_3	0.40	0.58	0.34	0.78	0.60	0.65	0.63	0.50	0.43	0.44	0.69	0.65	0.78	0.51	0.67	0.44	0.53	0.29
I_4	0.43	0.45	0.25	0.55	0.41	0.69	0.38	0.55	0.47	0.40	0.61	0.76	0.64	0.47	0.59	0.58	0.44	0.55
I_5	0.37	0.52	0.38	0.72	0.44	0.74	0.61	0.60	0.52	0.53	0.79	0.70	0.61	0.57	0.84	0.44	0.48	0.38
I_6	0.20	0.46	0.13	0.61	0.30	0.77	0.53	0.46	0.42	0.41	0.65	0.52	0.49	0.57	0.65	0.52	0.45	0.42
I_7	0.04	0.04	0.09	0.59	0.24	0.84	0.85	0.96	0.48	0.22	0.22	0.53	0.76	0.34	0.46	0.63	0.91	0.16
I_8	0.41	0.59	0.24	0.75	0.42	0.53	0.52	0.66	0.51	0.60	0.82	0.79	0.53	0.41	0.54	0.46	0.34	0.41
I_9	0.46	0.58	0.22	0.53	0.44	0.78	0.36	0.61	0.68	0.36	0.73	0.79	0.58	0.38	0.76	0.56	0.50	0.57
I_{10}	0.19	0.75	0.34	0.41	0.15	0.81	0.62	0.73	0.80	0.06	0.95	0.49	0.75	0.74	0.83	0.15	0.45	0.61
I_{11}	0.32	0.42	0.44	0.42	0.32	0.34	0.63	0.46	0.65	0.40	0.54	0.64	0.28	0.73	0.72	0.49	0.67	0.57
I_{12}	0.37	0.31	0.60	0.44	0.44	0.36	0.56	0.70	0.65	0.49	0.35	0.55	0.20	0.87	0.75	0.66	0.72	0.60
I_{13}	0.40	0.00	0.54	0.20	0.21	0.32	0.09	0.74	0.74	0.54	0.33	0.83	0.55	0.95	0.89	0.35	0.54	0.34
I_{14}	0.39	0.48	0.63	0.40	0.55	0.25	0.52	0.61	0.66	0.30	0.52	0.66	0.26	0.67	0.63	0.50	0.49	0.60
I_{15}	0.05	0.59	0.16	0.83	0.16	0.50	0.99	0.35	0.04	0.21	0.39	0.33	0.22	0.93	0.68	0.96	0.43	0.94
I_{16}	0.19	0.42	0.64	0.44	0.59	0.49	0.65	0.56	0.66	0.29	0.35	0.48	0.37	0.89	0.65	0.52	0.68	0.50
I_{17}	0.51	0.29	0.65	0.58	0.37	0.39	0.51	0.57	0.67	0.21	0.54	0.65	0.30	0.77	0.74	0.61	0.64	0.55
I_{18}	0.38	0.42	0.95	0.57	0.84	0.27	0.62	0.58	0.96	0.08	0.50	0.52	0.09	0.90	0.88	0.43	0.78	0.14
I_{19}	0.39	0.31	0.53	0.63	0.35	0.34	0.54	0.49	0.66	0.50	0.33	0.45	0.41	0.75	0.87	0.70	0.61	0.58
I_{20}	0.32	0.43	0.70	0.54	0.59	0.47	0.54	0.59	0.69	0.24	0.40	0.70	0.32	0.84	0.84	0.70	0.61	0.34
I_{21}	0.31	0.22	0.65	0.06	0.27	0.28	0.88	0.44	0.75	0.60	0.78	0.11	0.97	0.84	0.05	0.46	0.32	0.63
I_{22}	0.36	0.45	0.57	0.32	0.27	0.34	0.70	0.59	0.47	0.52	0.45	0.52	0.67	0.53	0.40	0.45	0.43	0.46
I_{23}	0.59	0.52	0.52	0.24	0.15	0.57	0.70	0.52	0.58	0.43	0.54	0.46	0.72	0.32	0.38	0.63	0.53	0.52
I_{24}	0.51	0.30	0.65	0.13	0.22	0.39	0.84	0.48	0.63	0.58	0.51	0.62	0.43	0.57	0.29	0.65	0.43	0.49
I_{25}	0.51	0.57	0.51	0.16	0.15	0.53	0.79	0.68	0.55	0.68	0.47	0.65	0.44	0.38	0.36	0.72	0.61	0.41
I_{26}	0.15	0.84	0.78	0.27	0.22	0.32	0.82	0.82	0.57	0.57	0.28	0.69	0.79	0.44	0.44	0.46	0.27	0.67
I_{27}	0.57	0.56	0.61	0.21	0.35	0.37	0.58	0.52	0.49	0.56	0.33	0.60	0.65	0.61	0.27	0.64	0.64	0.49
I_{28}	0.37	0.58	0.38	0.23	0.44	0.45	0.64	0.49	0.73	0.42	0.41	0.63	0.57	0.51	0.30	0.65	0.37	0.48
I_{29}	0.29	0.53	0.48	0.40	0.21	0.52	0.84	0.47	0.46	0.38	0.52	0.53	0.45	0.57	0.31	0.45	0.38	0.36
I_{30}	0.78	0.09	0.23	0.24	0.10	0.85	0.69	0.73	0.65	0.51	0.32	0.66	0.11	0.14	0.01	0.96	0.97	0.12

Figure 10: Initial Population Created for the Multi-objective Algorithm

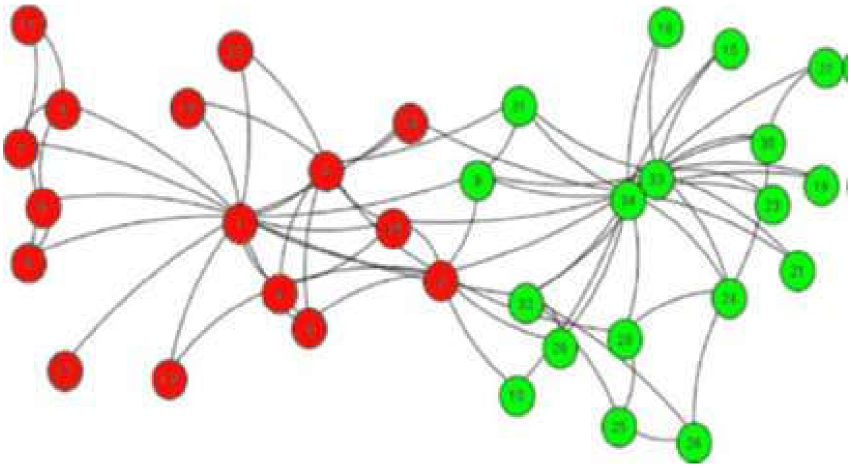


Figure 11: Communities found by POA for Zachary's Karate Club

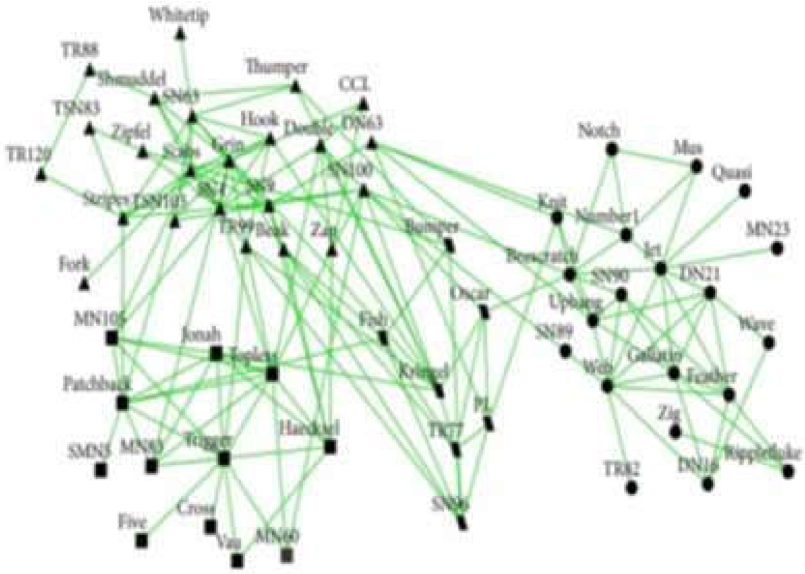


Figure 13: Communities found by POA for Dolphin Social Network

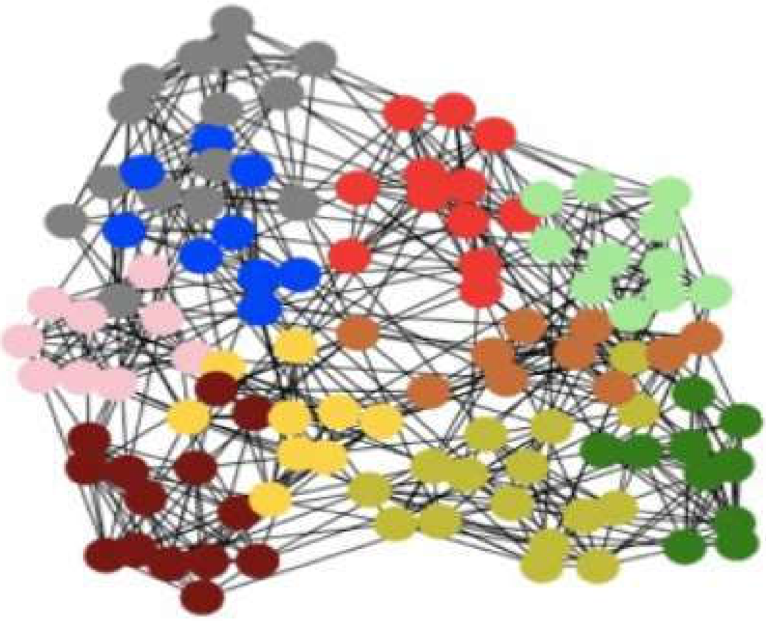


Figure 12: Communities found by POA for American College Football

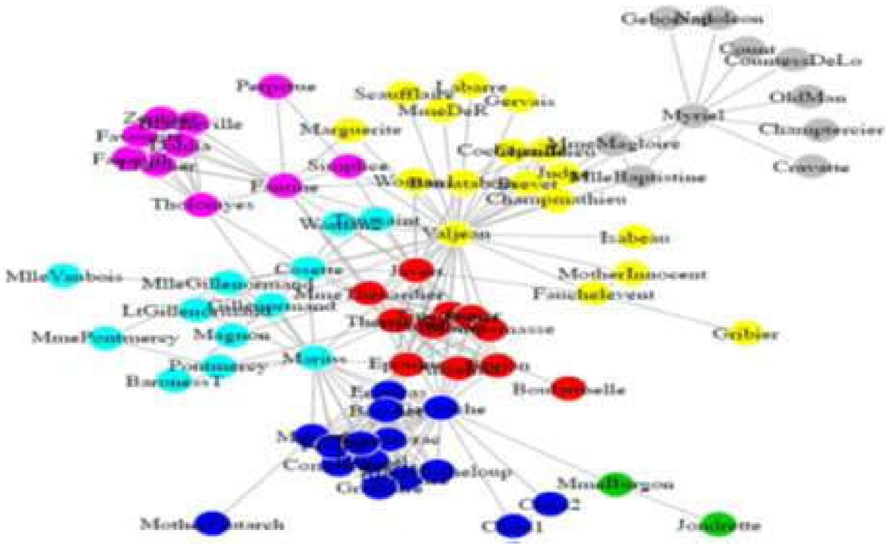


Figure 14: Communities found by POA for Lesmis

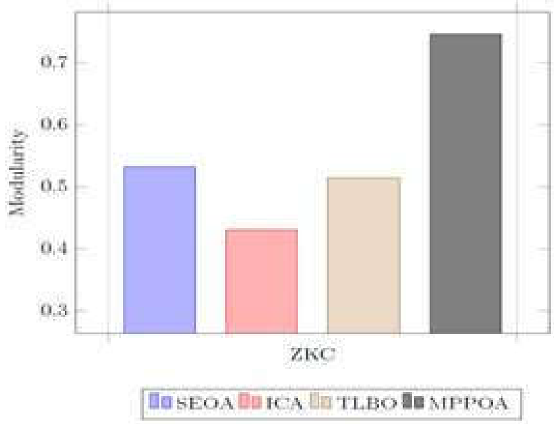


Figure 15: Modularity of Community Detection Over ZKC Data Set

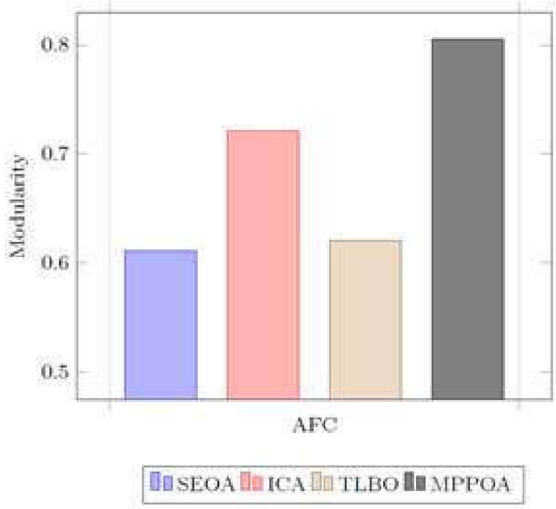


Figure 17: Modularity of Community Detection Over AFC Data Set

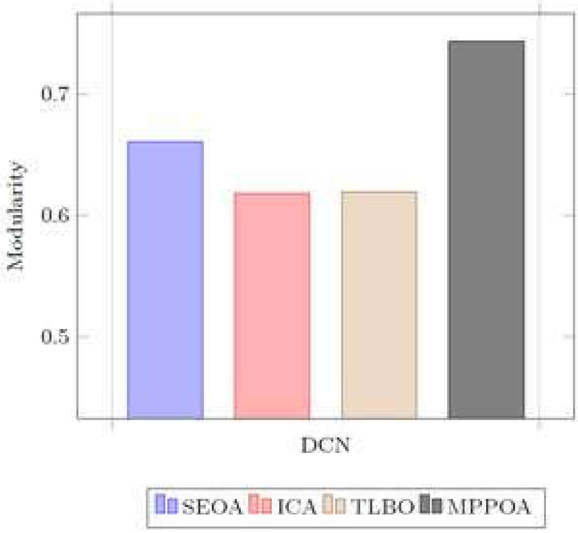


Figure 19: Modularity of Community Detection Over DCN Data Set

Where, x is the class label, c is the community structure, e is the Entropy and is the information gain for element for class label x. Performance evaluation of benchmark community detection algorithm with and without social theories are shown in Tables 19 and 20 as modularity and normalized mutual information, respectively. Both the evaluation parameter is significantly improved after incorporating social theories with community detection algorithm. The community detection algorithm Social- Emotional Optimization Algorithm (SEOA), Imperialist

Competitor Algorithm (ICA), Teaching Learning Based Optimization Algorithm (TLBO), MPPOA gain approximate 53.13%, 43.02%, 51.37%, 74.56% modularity and 81.02%, 86.51%, 86.24%, 90.87% NMI over ZKC data sets respectively, as shown Figure 15 and 16. MPPOA leads the modularity over the SEOA algorithm

and NMI information over ICA and TLBO algorithm, whose archive was best before MPPOA. Whereas over AFC data set, community detection algorithm SEOA, ICA, TLBO, MPPOA gain approximate 61.07%, 72.11%, 62.04%, 80.56% modularity and 82.33%, 63.24%, 78.23%, 87.63% NMI respectively, as shown Figure 17 and 18. MPPOA leads the modularity over the Whereas over DCN data set, community detection algorithm SEOA, ICA, TLBO, MPPOA gain approximate 66.10%, 61.79%, 61.96%, 74.35% modularity and 79.41%, 71.62%, 58.61%, 71.52%, 85.67% NMI respectively, as shown Figure 19 and 20.MPPOA leads the performance over the SEOA algorithm, whose archive was best before MPPOA. SEOA algorithm leads the modularity, whereas SBA and HSA algorithm achieves the highest NMI information. Whereas over BUP data set, community detection algorithm SEOA, ICA,

TLBO, MPPOA gain approximate 61.51%, 67.14%, 71.19%, 81.23% modularity and 69.55%, 57.12%, 52.51%, 76.52% NMI respectively, as shown Figure 21 and 22. MPPOA leads the modularity over the TLBO algorithm and NMI information over SEOA algorithm, whose archive was best before MPPOA. Whereas over LM data set, community detection algorithm SEOA, ICA, TLBO. MPPOA gain approximate 61.21%, 61.22%, 62.17%, 76.58% modularity and 69.55%, 57.12%, 52.51%, 74.59% NMI respectively, as shown Figure 23 and 24. MPPOA leads the modularity over the TLBO algorithm and NMI information over SEOA algorithm, whose archive was best before MPPOA.Whereas over WA data set, community detection algorithm SEOA, ICA, TLBO, MPPOA gain approximate 41.18%, 41.92%,

51.03%, 68.56% modularity and 58.22%, 42.68%, 51.22%, 71.25% NMI respectively, as shown Figure 25 and 26. MPPOA leads the modularity over the SEOA algorithm and NMI information over TLBO algorithm, whose archive was best before MPLM POA. The performance of Multi-Purpose Overlapping Community Detection With POA over social media data set varies with network density. It achieves a higher performance rate, higher dense ACF and ZKC network and relatively lower over lightly dense WA data set.

CONCLUSION

This paper proposed POA based single and multi-purpose function to discover overlapped community in social
networks. After the data representation was determined, the single-purpose algorithm was tested on artificial data with the program prepared in a python environment, and ensembles and overlapping nodes were tested. Using the same representation format for data, the multi-purpose algorithm was tested on artificial and actual world data. Costumes and overlapping nodes were determined with an approach that was not previously found in the literature. The proposed algorithm has been developed to optimise the modularity and internal density of social networks. The initial population of POA was created in a Python environment for the data used. The population formed was divided into a certain number of groups, and the power values of each group were calculated. While strong groups show joining according to the determined combination probability value, the vulnerable groups are eliminated from the population according to the determined deletion probability. At the same time, this paper presents a comparative analysis of proposed MPPOA with three meta-heuristic overlapping community detection algorithms over six different social media-based data sets. This paper observed that community detection algorithm SEOA, ICA, TLBO over social media data set varying with network density and its achieves higher performance rate higher dense ACF and ZKC network and relatively lower over lightly dense WA data set. Moreover, performance of MPPOA not to much varying network density. Its extract higher informative community over the higher dense network as compare to TLBO and at the same time, gain better results with the lower dense networks as compare to SEOA algorithm.

References

[1] K. Kadkhoda Mohammadmosaferi and H. Naderi, “Afif: Automatically finding important features in community evolution prediction for dynamic social networks,” Computer Communications, vol. 176, pp. 66–80, 2021.

[2] H. A. Deylami and M. Asadpour, “Link prediction in social networks using hierarchical community detection,” in 2015 7th Conference on Information and Knowledge Technology (IKT), pp. 1–5, 2015.

[3] F. Karimi, S. Lotfi, and H. Izadkhah, “Communityguided link prediction in multiplex networks,” Journal of Informetrics, vol. 15, no. 4, p. 101178, 2021.

[4] A. Fang, “The influence of communication structure on opinion dynamics in social networks with multiple true states,” Applied Mathematics and Computation, vol. 406, p. 126262, 2021.

[5] W. Liu, T. Suzumura, L. Chen, and G. Hu, “A generalized incremental bottom-up community detection framework for highly dynamic graphs,”

in 2017 IEEE International Conference on Big Data (Big Data), pp. 3342–3351, 2017.
[6] Q. Chen and L. Wei, “Overlapping community detection of complex network: A survey,” in 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), pp. 513–516, 2019.
[7] D. K. Singh, R. A. Haraty, N. C. Debnath, and P. Choudhury, “An analysis of the dynamic community detection algorithms in complex networks,” in 2020 IEEE International Conference on Industrial Technology (ICIT), pp. 989–994, 2020.
[8] K. Fatania, D. Joshi, T. Patalia, and Y. Jejani, “A comparison of overlapping community detection in large complex network,” in 2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC), pp. 1–6, 2019.
[9] B. Shen, N. Wang, and H. Qiu, “A new genetic algorithm for overlapping community detection,” in 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 766–769, 2014.
[10] C. He, H. Liu, Y. Tang, S. Liu, X. Fei, Q. Cheng, and H. Li, “Similarity preserving overlapping community detection in signed networks,” Future Generation Computer Systems, vol. 116, pp. 275–290, 2021.
[11] Y. Yan and H. Xie, “Detection of hierarchical overlapping communities in complex networks,” in 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS), pp. 893–896, 2019.
[12] D. DEMIROL, F. OZTEMIZ, and A. KARCI, “Performance comparison of physics based meta-heuristic optimization algorithms,” in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), pp. 1–5, 2018.
[13] L. de Marcos, A. Garc´ıa, E. Garc´ıa, J. J. Mart´ınez, J. A. Guti´errez, R. Barchino, J. M. Guti´errez, J. R. Hilera, and S. Ot´on, “An adaptation of the parliamentary metaheuristic for permutation constraint satisfaction,” in IEEE Congress on Evolutionary Computation, pp. 1–8, 2010.
[14] X. Zhou, Y. Liu, J. Wang, and C. Li, “A density based link clustering algorithm for overlapping community detection in networks,” Physica A: Statistical Mechanics and its Applications, vol. 486, pp. 65–78, 2017.
[15] Q. Askari, I. Younas, and M. Saeed, “Political optimizer: A novel socio-inspired meta-heuristic for global optimization,” Knowledge-Based Systems, vol. 195, p. 105709, 2020.
[16] E. Osaba, J. Del Ser, D. Camacho, M. N. Bilbao, and X.-S. Yang, “Community detection in networks using bioinspired optimization: Latest developments, new results and perspectives with a selection of recent metaheuristics,” Applied Soft Computing, vol. 87, p. 106010, 2020.
[17] X. Zhu, Y. Ma, and Z. Liu, “A novel evolutionary algorithm on communities detection in signed networks,” Physica A: Statistical Mechanics and its Applications, vol. 503, pp. 938–946, 2018.

[18] X. Wan, X. Zuo, and F. Song, “Solving dynamic overlapping community detection problem by a multiobjective evolutionary algorithm based on decomposition,” Swarm and Evolutionary Computation, vol. 54, p. 100668, 2020.
[19] W. W. Zachary, “An information flow model for conflict and fission in small groups,” Journal of Anthropological Research, vol. 33, no. 4, pp. 452–473, 1977.
[20] D. Lusseau, “The emergent properties of a dolphin social network,” Proceedings of the Royal Society of London. Series B: Biological Sciences, vol. 270, no. suppl 2, pp. S186–S188, 2003.
[21] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” Proceedings of the National Academy of Sciences, vol. 99, pp. 7821–7826, 2002.
[22] R. A. Rossi and N. K. Ahmed, “The network data repository with interactive graph analytics and visualization,” in AAAI, 2015.
[23] M. E. J. Newman, “Modularity and community structure in networks,” Proceedings of the National Academy of Sciences, vol. 103, no. 23, pp. 8577–8582, 2006.

“Social Media Community Detection using Machine Learning-based Clustering Analysis “

A

Thesis

Submitted towards the Requirement for the Award of Degree of

Doctor of Philosophy

In

COMPUTER SCIENCE ENGINEERING

Under the Faculty of Engineering and Technology

By

RAVINDRA SINGH YADAV

(Enrollment No- 161588517256)

Under the Supervision of

Dr. BALVEER SINGH

Professor

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING &IT

Year - 2024

P.K. University Shivpuri M.P. -473665

Conclusions

End users' engagement in corporate marketing, political propaganda, educational activities, entertainment, and commercial activity social media has significantly risen in social media era. End users' inclination foundation has been facilitated by social media.

participation by forming groups of people who have same interests and viewpoints on a range of topics, including local global both politics, products, and concerns. Social media behavior, including likes, dislikes, shares, comments, friends, and profile information, can shape end users' preferences. In order to tackle the delicate character of community structure social media, this thesis focuses on overlapping community finding and the decreased engagement of passive like-minded individuals.

This study solves community discovery that coincides a method has not used previously, and first presents a thorough assessment on overlapping community structure social networks, is commonly experienced in everyday life. It have been shown via investigations and analyses that the methods used to identify overlapping communities in the social networks offer answers to this issue by focusing on single goal. This article also provides a comparative examination of six distinct social media based sets using meta-heuristic overlap community recognition techniques.

community detection algo SEOA,BSO, , HSA, GLOA,HGFA, SBA, ICA, and TLBO were examined in this section in relation to social media sets. The results showed these algorithms achieved greater performance rates in dense ACF networks and comparatively lower rates in lightly dense a WA data sets. Additionally, across the greater dense network, TLBO , ICA are able to extract a higher informative community. The semantic relation-based modularity-optimized community recognition approach for heterogeneous networks is then presented in this paper. This part attempts to use content analysis the network try to raise the modularity ratio of the jointly link analysis and the network. As a result, network was constructed using the similarity values individuals' shares as indirect linkages.

Additionally, this study introduced content- and link-based approaches for the greedy hierarchical clustering algorithm, which makes advantage of indirect connections the network and structure to guarantee that nodes that are closest to one another are prioritized in both topological and semantic groups. The influence semantic relations on optimization algorithms, namely Parliamentary Optimization (POA) and Modularity Optimization (MOA), community discovery is compared in this section. Ultimately, six real network based heterogeneous network information sets were used to test modularity and NMI the study that was given, and the results showed a sufficient modularity rate across the resulting informative community.

Ultimately, this study develops a Parliamentary Optimization (MPPOA) community identification approach based on Multi Purpose functionality that extracts user space for the purpose influencing passive last users. In addition, the semantic relation-based modularity-optimized community recognition technique for heterogeneous networks is presented in this thesis. By combining the network's content analysis with link analysis, this thesis seeks to improve the network's modularity value. As a result, similarity values individuals' shares were computed and added as indirect linkages to the network. Significantly, MPPOA organized a very interesting and dense community. obtained 94% NMI and around 93% modularity throughout the community structure across several data set variations. By comparison, MPPOA outperforms best acquired result benchmark approach with the combination of graphical social theory over 6six distinct actual network data sets by 2.24% improvement in modularity and 3.57% improvement in NMI, respectively.