Deepali Piple, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Analysis of Modularity Based Community Detection Algorithms to Detect Communities in Social Network Analysis

Deepali Piple[1], Mukesh Sakle[2], Shaligram Prajapat[3]
[1]Research Scholar, Department of information Technology, Shri G.S. Institute of Technology and Science, Indore, India
[2]Assistant Professor, Department of information Technology, Shri G.S. Institute of Technology and Science, Indore, India
[3]Professor, IIPS, Devi Ahilya Vishwavidhyalaya, Indore, India.

Abstract- Social Network Analysis relies upon community detection to identify groups of nodes within a network that maintain stronger inter-node relationships than their node linkages to all other components of the network. Community Detection enables researchers to understand how complex networks construct and establish their functional connections. This research tests various community detection techniques by way of several modularity-based community detection techniques using established benchmark social networks. The study evaluates four community detection algorithms: Label Propagation Algorithm (LPA) and Clauset-Newman-Moore (CNM) and Louvain and Leiden.Each of the four community detection algorithms was evaluated against five standard benchmark social networks including Karate, Dolphin, Football, Facebook, Polbooks, Les Misérables and Jazz.Performance evaluation metrics used for each of the four algorithms included the Modularity Index (MI), Conductance, Normalized Mutual Information (NMI) and the Adjusted Rand Index (ARI). Results from this research demonstrated that the Louvain method consistently produced higher MI values and had consistent performance across all of the various test data sets, while the LPA method performed all other methods in networks where community structures are clearly visible. The research also found that modularity-based optimizations successfully identified critical social network communities.

Keywords: Community, Community Detection, Modularity, NMI, Social Network Analysis.

I. INTRODUCTION

Social networks are an essential resource which researchers seek to analyse complex behaviours and interactions between individuals, organizations, and other entities. The networks exist as graph where entities (individual) are represented by nodes and the relationships between them are shown through edges. Community Detection is a critical problem in the area of social network analysis, as the goal of this approach is to discover groups of nodes that are connected by a greater number of edges than the connections to the remaining nodes in the network. Identifying such communities can help to expose the underlying structures, facilitate understanding of social behaviours, and provide insights into meaningful patterns or trends in large-scale networks. In recent years, numerous community detection methods have emerged for analysing the organizational structure of networks.

Among these methods, modularity-based approaches are gaining significant attention, as they are capable of measuring the quality of the community being identified within a network. Modularity measuresthe number of edges link nodes betweenlocal entities which exist within the same community compared to the edges that connect different communities, and higher modularity values specify stronger community structures. The Clauset–Newman–Moore (CNM) algorithm, Louvain method, Leiden algorithm and several algorithms have been developed to optimize modularity. The Label Propagation Algorithm (LPA) allows an efficient method for community detection which uses its scalable label diffusion-based detection method.

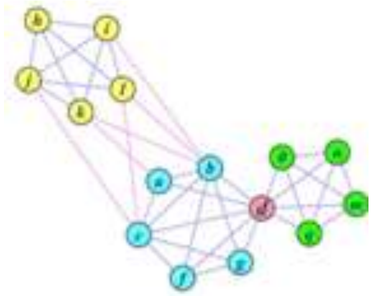


Figure 1: Community Detection Network

The effectiveness of numerous community detection algorithms will depend on both network size and structural design together with existing connection patterns. The evaluation process on multiple datasets requires testing different algorithms because this method helps researchers evaluate which algorithms successfully detect significant communities. The study examines various community detection algorithms through LPA, CNM, Louvain, and Leidenby using benchmark social network datasets. The algorithms are evaluated with metrics such as modularity, conductance, normalized mutual information (NMI), and adjusted rand index (ARI).

The main objective of this study is to compare the effectiveness of modularity-based community detection algorithms on different social network datasets and analyse their performance in terms of community quality and accuracy. The results provide insights into the strengths and limitations of each algorithm and highlight the importance of modularity-based optimization techniques for social network analysis.

II. LITERATURE REVIEW

The initial research on community detection used graph-theoretic methods to find structural boundaries in complex networks. In 2002, Girvan and Newman developed the first community detection algorithm which uses edge betweenness centrality as its basis. Their method uses an iterative process to eliminate edges which possess the highest betweenness values until the system shows its community structures. The algorithm successfully found communities in both artificial networks and actual networks which established itself as a key reference point for community detection studies [1].

Newman et al introduced their evaluation method for community structures through the implementation of their modularity metric. The modularity metric measures the actual number of edges that exist within communities against the predicted number of edges which would exist in a random network. Community detection algorithms adopted this metric because it enabled researchers to assess various network partitions while serving as an effective objective function [2]. They developed modularity as a community structure evaluation metric which has become an essential quality measurement in their research. The modularity metric compares actual community edge connections to the expected edge connections in a random network. The strength of community structures increases with ascending modularity measurements. Researchers used this metric to assess community partitions while optimizing community detection algorithms which developed from this metric.

Clauset et al explained scalability problems through their development of a greedy algorithm that enables modularity optimization in extensive networks. The method executes community merging operations to identify the maximum modularity enhancement which will occur during each merging operation. The algorithm achieves substantial computational work reductions through its use of data structures which enable community detection in networks that have more than 1,00,000 nodes. The research showed that researchers could use modularity-based methods to analyze extensive networks [3].

Pons and Latapy presented a new method through their creation of the Walktrap algorithm which uses random walk processes for its implementation. The central idea is that short random walks will stay within areas of a network which have high connection density. The algorithm uses probability distribution analysis of random walks to determine node similarity which it uses to create communities through hierarchical clustering. The approach achieves an optimal combination of computational performance and accurate detection results [4].

Modularity-based methods have achieved successful outcomes, but they encounter various challenges. The resolution limit problem arises from Fortunato and Barthélemy established that modularity optimization fails to identify small communities within extensive networks. Their analysis exhibited that communities below a specific size threshold could be consolidated into extensive entities, despite their well-defined structural characteristics. After this work, Researchers started looking into different methods to improve, and which directed to the development of better community detection systems [5].

Reichardt and Bornholdt recognized a statistical mechanics framework that helps researchers to find communities in complex network. The researchers utilized the spin-glass model to formulate their solution, which conceptualizes network nodes as spins and their communities act as different spin states. The algorithm finds the best optimal community partitions by minimizing its process of system energy. The method formed a connection between statistical physics and network analysis, which provided researchers a way for understanding community detection [6].

The Map Equation method uses information theory developed by Rosvall and Bergstrom as its foundation. The method uses random walks to model information flow through networks while it detects network communities by reducing the random walk description length of its trajectory. The algorithm uses network partitioning to create modules which help identify important community patterns through efficient information flow compression. The method has been extensively used to analyze large networks which include citation networks and biological networks [7].

The Louvain method developed by Blondel et al. in 2008 patent a fundamental change in community detection methods that operate on large-scale networks. The algorithm executes modularity optimization through its two-phase process which combines local node movement with community aggregation. The Louvain algorithm is one of the highly efficient ways to analyze networks becauseit is very fast and capable of detecting communities which containing millions of nodes, making it one of the most widely used methods in network analysis [8].

While they were working on new algorithms, researchers looked at evaluation methods together with benchmark datasets. The LFR benchmark model initiated by Lancichinetti, Fortunato, and Radicchi uses power-law distributions to start with synthetic networks which generate both node degrees and community sizes. The benchmark is a better way to illustrate how different real-world network, and it became the standard way for testing community detection algorithms [9].

Arenas et al. developed a multi-resolution modularity frame work to solve problems that come up with fixed-resolution methods. Their method has a controllable parameter which enables detection of communities at different levels of size. This system helps users to find both main and sub-ordinate groups in a single network and it resolves issues which earlier modularity-based approaches encountered according to the reference [10].

The comprehensive survey conducted by Fortunato evaluated all present community detection technique and summed up their theoretical foundations, algorithms, and challenges in the field. The research categorized community detection techniques into numerous classifications, which incorporate graph partitioning, modularity optimization, spectral clustering and statistical inference approaches. This study has become a key source of information for researchers working in network science [11].

The study on modularity optimization revealed new difficulties that are required to solved. Montjoye et al. showed that modularity maximization outcomes in several community partitions which can compare similar modularity measurements. The research found that determining the optimal community structure in complex networks appears as a difficult task [12].

A consensus clustering framework which combines several community detection results to found a consensus matrix that helps to improves system reliability was anticipated by Lancichinetti et al. By combining data from various algorithm tests, and the method becomes more essential for community distribution that is crucial for handling large and dynamic networks [13].

Yang and Leskovec enhanced the evaluation of community detection method which examines by analyzing networks that contained true community structures. In order to evaluate various structural metrics that help in identifying significant community patterns and that study observed real networks. According to the study, conductance measurements are better than modularity measurements at identifying true communities [14].
Shi et al. used a multi-objective evolutionary approach to create a community detection system that applies multiple objectives. Their method addresses multiple objectives that deals how strongly community members connect relate to one another and to members from different communities. To produce the framework Pareto-optimal community partitions that help to reveal complete network structure by using evolutionary algorithms [15].

Recent research has focused on enhancing the scalability and dependability of community detection algorithms. Traag et al. developed the Leiden algorithm acts as an enhanced version of the Louvain method. The Leiden algorithm delivers well-connected communities, which enable faster and more accurate community detection results [16].

Scientists have developed new methods which combine optimization techniques with machine learning methods to improve community detection results. Researchers have used evolutionary algorithms together with metaheuristic software to enhance modularity optimization results for complex network systems. Researchers developed iterative greedy algorithms which improve the efficiency of local search operations. The researchers have started using deep learning techniques to observe network structures in their currentresearch. Graph Neural Networks (GNNs) analyze node structural patterns to identify communities and predict links between them. The methods combine community detection results with neural network models to enhance prediction accuracy in extensive network systems [17].

The current research determines that community detection methods evolved from their original graph-based algorithms to modern optimization techniques and deep learning methods. Researchers need to investigate hybrid and multi-objective community detection methods because current techniques struggle to identify communities in large-scale dynamic networks that have overlapping structures.

III. METHODOLOGY

The study shows a community detection framework for social networks which uses modularity-based detection methods. The methodology comprises a systematic workflow that involves five stages: data acquisition, preprocessing, graph creation, community detection via modularity optimization, and performance evaluation. The first step requires researchers to find an appropriate social network dataset for their work. The dataset represents each entity in the network as a node while edges show the connections between these nodes. The dataset can be obtained from network repositories which provide public access to their data or from common benchmark datasets which researchers use in their community detection research. The datasets show how network elements connect with each other which helps researchers build their analysis framework.

After the dataset collection process, the network data enters a preprocessing stage which starts data consistency and reliability. The process includes three main tasks: duplicate edge elimination, missing or inconsistent value resolution, and dataset conversion into a format which enables network analysis. The preprocessing step helps to create a dataset which accurately shows node relationships while take away any interference that could lead to incorrect community detection results.

After the cleaning process of the dataset, the network transitions come into a graph representation. The social network represents as a graph which consists of nodes (V) and edges (E) that link these nodes. In this illustration, nodes correspond to users or entities in the network, while edges indicate the interactions or relationships between them these entities. Graph construction provides the structural foundation required for applying community detection algorithms.

After constructing the graph, researchers use modularity-based optimization techniques to find communities within the graph. Modularity serves as a common metric which measures network community division strength through the comparison of actual edge connections within communities to the expected edge connections of a random network. The most valuable network partitioning exists at the point where researchers achieve maximum modularity value (Q) assessment. The Clauset–Newman–Moore (CNM) algorithm together with the Louvain algorithm and the Leiden algorithm enable this phase to use its algorithms for community detection through iterative node grouping. The algorithms operate by distributing nodes among different communities while they continue to modify the community structure until their modularity score reaches its peak point.

The network community structure becomes visible through the partitioning process which follows modularity optimization. The process in which each community contains nodes that have stronger connections to each other than to members of other communities, discovers the basic structure of the social network, which enables researchers to discover groups of users or entities that have strong relationships with each other.

Performance metrics evaluate the identified communities to determine the results quality. The primary evaluation metric is modularity (Q), which measures how well the detected communities represent the structure of the network. Accuracy and Normalized Mutual Information (NMI) serve as additional evaluation measures which require ground-truth community information for assessment. The metrics present numerical measurements which demonstrate how well the community detection method operates.

The last step of the process involves researchers to study and display their uncovered communities to determine the network's basic structure. To reveal the node of network grouping patterns, the analysis observes community sizes and connectivity patterns together with modularity scores. The study presents valuable information about social network structure which demonstrates the success of using modularity-based community detection methods.

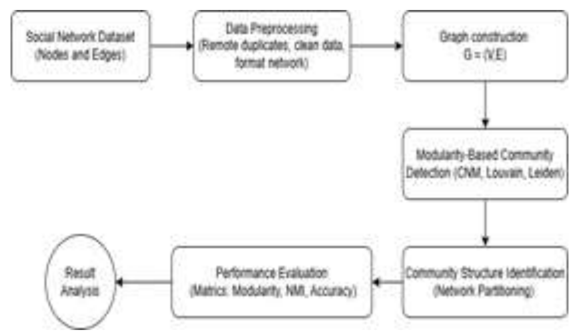


Figure 2: Workflow of Modularity-Based Community Detection in Social Networks

IV. RESULTS AND DISCUSSION

This section shows the performance of community detection algorithms: Label Propagation Algorithm (LPA), Clauset–Newman–Moore (CNM), Louvain, and Leiden across social network datasets including Karate, Dolphin, Football, Facebook, Polbooks, Les Misérables, and Jazz. The performance was evaluated using Modularity Index (MI), Conductance, Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI), which measure the strength of community structures, separation between communities, and similarity with ground-truth communities.

Table 1: Dataset Description

Dataset	|V|	|E|	Description
Karate	34	78	Social network of friendships between 34 members of a karate club at a US university in the 1970s
Dolphine	62	159	An undirected social network of frequent associations between 62

			dolphins in a community living off Doubtful Sound, New Zealand.
Football	115	613	Network of American football games between Division IA colleges during regular season Fall 2000.
Facebook	4039	88234	Social circles (or friend lists) from facebook
Pollbook	104	441	Networks of books about US politics
Les-mis	77	254	Coappearance network of characters in the novel Les Misérables.
Jazz	198	2742	Musician networks

The performance assessment of detected communities employed four evaluation metrics which included modularity, conductance, Normalized Mutual Information (NMI), and Adjust Rand Index (ARI). The strength of community structure in a network is evaluated by modularity which compares internal edge density between communities to the edge density of a simulated random network. Conductance measures the fraction of edges connecting a community to the rest of the network, with lower values indicating better community separation. The value of NMI closer to 1 indicate better agreement, and itassesses the similarity between the detected communities and the ground-truth community structure. The higher value of ARI metric values indicates better performance in identifying community structures, and it compares two clustering results by measuring their similarity while accounting for random chance.

Table 2: Comparative performance of different community detection algorithms on social network datasets

Dataset	Algorithms	MI	Conductance	NMI	ARI	Comm
Karate	LPA	0.436602	0.297024	0.505627	0.45878	4
	CNM	0.410965	0.28083	0.564607	0.568439	3
	Louvain	0.434521	0.202778	0.691249	0.684142	3
	Leiden	0.225764	0.598148	0.129088	0.117299	4
Dolphin	LPA	0.515427	0.29327	0.607702	0.419761	5
	CNM	0.495491	0.316318	0.557135	0.46586	4
	Louvain	0.523338	0.296535	0.484342	0.347197	5
	Leiden	-0.02102	0.833678	0.015791	0.00231	5
Football	LPA	0.580713	0.367334	0.91015	0.851441	13
	CNM	0.549741	0.277871	0.697732	0.474098	6
	Louvain	0.604184	0.281662	0.850554	0.704085	9
	Leiden	0.013886	0.887326	0.238894	0.006313	10
Facebook	LPA	0.817509	0.257346	0.044505	0.0019985	71
	CNM	0.777378	0.054633	0.013553	0.001515	13
	Louvain	0.834906	0.054215	0.014355	0.0023588	16
	Leiden	0.23482	0.759889	0.018804	0.0023974	17
Pollbook	LPA	0.514158	0.308903	-	-	6
	CNM	0.501974	0.254911	-	-	6
	Louvain	0.52662	0.239248	-	-	5
	Leiden	0.318764	0.509128	-	-	5
Les-mis	LPA	0.337664	0.265766	-	-	7
	CNM	0.472942	0.287746	-	-	5
	Louvain	0.565822	0.217492	-	-	6
	Leiden	0.132837	0.65277	-	-	6
Jazz	LPA	0.281597	0.16524	-	-	3
	CNM	0.438908	0.366359	-	-	4
	Louvain	0.440839	0.318477	-	-	4
	Leiden	0.118682	0.634943	-	-	4

Table 2 illustrate the result of datasets. The Louvain algorithm achieved the highest modularity along with the highest NMI and ARI for the Karate dataset, indicating strong agreement with the known community structure. CNM also demonstrated competitive performance, and Leiden produced lower evaluation scores. In the dolphin dataset, Louvain had the highest modularity, while LPA achieved the best NMI value. But, ARI values for the different algorithms were only moderate, which means that they only partially agreed with the actual communities.

For the Football network, which has well-defined communities, LPA attained the highest NMI and ARI, showing that effectiveness at detecting clear community structures, while Louvain also well performed with the highest modularity score. Louvain achieved the highest modularity in the larger Facebook network, which indicates that the strong structure was very well partitioned. However, all of the algorithms had low NMI and ARI values because there was the absence of clear ground-truth communities. For the Polbooks and Les Misérables datasets, Louvain once again produced the highest modularity values and along with relatively low conductance, which shows that the communities are well-separated. Similarly, in the jazz network, Louvain and CNM also had strong modularity scores, but Louvain had better conductance values.

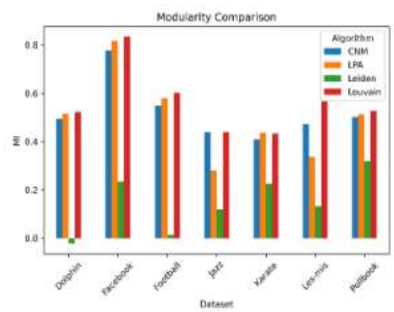
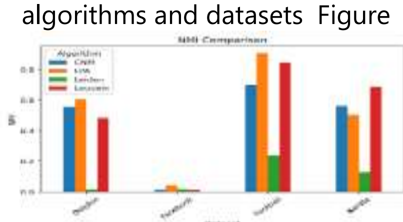


Figure 3: Modularity (MI) comparison across algorithms and datasets Figure



4: Conductance comparison across algorithms and datasets

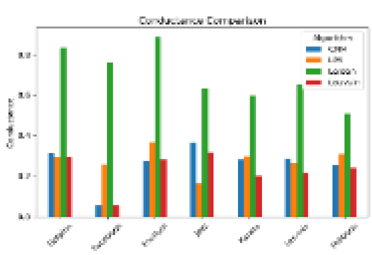
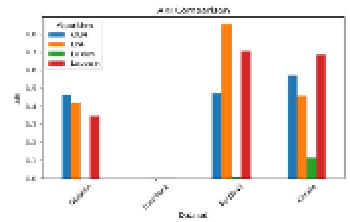


Figure 5: NMI comparison across algorithms and datasets Figure



6: ARI comparison across algorithms and datasets

The results demonstrate that the Louvain algorithm consistently produces overall high modularity values and stable performance across utmost datasets. The ground-truth defined datasets, in which LPA performs well in networks and Louvain provides a better balance between community quality and stability across networks of different sizes. Therefore, Louvain appears to be the most effective method among the evaluated algorithms for modularity-based community detection in social networks.

V. CONCLUSION AND FUTURE WORK

Community detection functions as an essential method which helps researcher's study how social networks create their structural frameworks through their node connections that display intense internal relationship patterns. The research team conducted their assessment of community detection algorithms through their testing of Label Propagation Algorithm (LPA) and Clauset–Newman–Moore (CNM) and Louvain and Leiden algorithms on multiple benchmark social network datasets which included Karate and Dolphin and Football and Facebook and Polbooks and Les Misérables and Jazz networks. The evaluation process used evaluation metrics which included modularity and conductance and normalized mutual information (NMI) and adjusted rand index (ARI) to assess the performance of these algorithms.

The results of the experiment indicate that different algorithms accomplish differently depending on the network's characteristics. Higher modularity values achieved by the Louvain algorithm among most datasets and it performed well while finding strong community structures. The LPA algorithm performed well in networks with clear ground-truth communities, but the CNM algorithm provided balanced results but generally found fewer communities overall. The Leiden algorithm should expand partitions better, but it resulted in lower performance for several datasets in this study.

The research findings indicate that social networks use modularity-based optimization techniques, including the Louvain algorithm as a reliable and effective approach for community detection. Future research may motivate on developing hybrid or enhanced modularity-based methodologies capable of handling extensive large-scale networks and overlapping community structures with greater effectiveness.

REFERENCE

1. Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." Proceedings of the national academy of sciences 99, no. 12 (2002): 7821-7826.
2. Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." Physical review E 69, no. 2 (2004): 026113.
3. Clauset, Aaron, Mark EJ Newman, and Cristopher Moore. "Finding community structure in very large networks." Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 70, no. 6 (2004): 066111.
4. Pons, Pascal, and Matthieu Latapy. "Computing communities in large networks using random walks." In International symposium on computer and information sciences, pp. 284-293. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.
5. Fortunato, Santo, and Marc Barthelemy. "Resolution limit in community detection." Proceedings of the national academy of sciences 104, no. 1 (2007): 36-41.
6. Reichardt, Jörg, and Stefan Bornholdt. "Statistical mechanics of community detection." Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 74, no. 1 (2006): 016110.
7. Rosvall, Martin, and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure." Proceedings of the national academy of sciences 105, no. 4 (2008): 1118-1123.
8. Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast unfolding of communities in large networks." Journal of statistical mechanics: theory and experiment 2008, no. 10 (2008): P10008.
9. Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi. "Benchmark graphs for testing community detection algorithms." Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 78, no. 4 (2008): 046110.
10. Arenas, Alex, Alberto Fernandez, and Sergio Gomez. "Analysis of the structure of complex networks at different resolution levels." New journal of physics 10, no. 5 (2008): 053039.
11. Fortunato, Santo. "Community detection in graphs." Physics reports 486, no. 3-5 (2010): 75-174.
12. Good, Benjamin H., Y. de Montjoye, and Aaron Clauset. "The performance of modularity maximization in practical contexts, eprint." arXiv preprint arXiv:0910.0165 5 (2009).
13. Lancichinetti, Andrea, and Santo Fortunato. "Consensus clustering in complex networks." Scientific reports 2, no. 1 (2012): 336.
14. Yang, Jaewon, and Jure Leskovec. "Defining and evaluating network communities based on ground-truth." In Proceedings of the ACM SIGKDD workshop on mining data semantics, pp. 1-8. 2012.
15. Shi, Chuan, Zhenyu Yan, Yanan Cai, and Bin Wu. "Multi-objective community detection in complex networks." Applied Soft Computing 12, no. 2 (2012): 850-859.
16. Traag, Vincent A., Ludo Waltman, and Nees Jan Van Eck. "From Louvain to Leiden: guaranteeing well-connected communities." Scientific reports 9, no. 1 (2019): 5233.
17. Liu, Chunjiang, Yikun Han, Haiyun Xu, Shihan Yang, Kaidi Wang, and Yongye Su. "A community detection and graph-neural-network-based link prediction approach for scientific literature." Mathematics 12, no. 3 (2024): 369.