A Survey On Community Identification in Dynamic Network

Oumaima IZEM, Youssef ZAZ, Ali YOUNES,
Nassime El RHARBI

*CSSE Labs, Faculty of scineces Tetouan, Av. Sebta , Tetouan, 93002,Tanger-Tetouan-Al Hoceima Morocco. Contributing authors:

oumaima.izem@etu.uae.ac.ma;
youssef.zaz@uae.ac.ma; ayounes@uae.ac.ma; nassime.elrharbi@etu.uae.ac.ma;

Abstract

Identifying community in dynamic networks is a process of determining the community structure of a continuously changing network , in order to provide significant insight into its characteristics and its functionality, allowing for the exact purpose of understanding deeply the operating and crucial features of current systems and investigating their underlying processes. Static networks have been the main focus of community detection research . However , as time went on, the research drifted to more scalable networks known as dynamic networks, which is the main study in our journal, since this type of network has been in multiple crucial applications such as social media, security, and public health. Accord - ingly, our research will be based on recent algorithms and community detection techniques, since, as we will it proved rather be a difficult challenge, along with diving into the brains behind each algorithm, comparing the effectiveness of each process there is as of today.

Keywords: Community detection, Dynamic Networks, Community Detection Algorithms

1 Introduction

Network science is a field that studies the way to determine network behavior by analyzing them to uncover the principles that structure these dynamic networks using

M. Khaldi et al. (eds.), *Proceedings of the E-Learning and Smart Engineering Systems (ELSES 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 14,
https://doi.org/10.2991/978-94-6463-360-3_34

mathematical theories. This study has applications in fields such as physics and computer science and much more. Many systems take the form of a network, and it is often represented by graphs, where the network is defined as a collection of objects that are connected to one another. These connections can be represented by transforming the objects into nodes, and the connections between nodes into edges. The majority of community detection research focuses on static networks. Static networks are networks that do not evolve over time and have a fixed structure that remains the same. However, a lot of networks, including biological, social, and transportation networks, are dynamic and constantly changing. Therefore, community detection algorithms may not operate as effectively in dynamic networks as they do in static ones. Furthermore, real-world networks evolve over time, and they further display a certain dynamism. In particular, it is possible to occur some groups of nodes have similar properties or functions, that tend to form highly cohesive subgraphs and have connections between them that are denser than the connections with the rest of the network, which are called communities.

Community detection is the process of dividing communities in each snapshot of a continuously changing network, which can help in understanding the underlying structure of dynamic networks, and provide insights into the mechanisms of its evolution. However, community detection in dynamic networks, received a significant amount of attention, since it remains a challenge in network analysis, due to its evolving structure as well as it requires scalable and efficient methods to detect communities, as the complexity of networks can be extensive, It is thus not surprising that discovering communities has been broadly investigated over the last few years. As a result, several algorithms have been come up to identify communities in dynamic networks, which are divided into two principal lines of research: graph partitioning, which involves dividing a graph into smaller subgraphs called partitions by reducing the number of cross-edges between the partitions, making sure that each section is approximately the same size. It has numerous applications in domains including computer sciences, computing, and integrated circuit design. Along with, clustering or community structure detection that group nodes into clusters based on similarity or connectivity patterns. Clustering deals with multiple attribute types, on the other hand, community identification is specifically for network analysis which depends on a single attribute type called edge. They have various applications, including social network analysis biology. It is worth mentioning that these two pieces of research do indeed address the same issue. Nevertheless, community detection algorithms have gained more attention over the last few years, as they provide a meaningful interpretation of the network's structure and function. These algorithms are particularly useful for analyzing dynamic networks where nodes and edges may change over time.

2 Related Work

This section reviews some of the most prominent theories and research papers that analyze dynamic community detection algorithms. Our research methodology involved a comprehensive search of academic databases such as IEEE, SpringerLink, ScienceDirect, and Clarivate to identify the core concepts that researchers have used to compare these algorithms. In addition, researching keywords and field specification was our first concern, along with community detection algorithms including the Louvain algorithm, Label Propagation algorithm, Girvan-Newman algorithm, and Propagation algorithm. After this brief introduction, we will evaluate the quality of the articles that have been studied up to this point. [1]

One of the most frequently used methods for detecting communities in dynamic networks is the Louvain algorithm developed by Blondel et al. [5] The Louvain is based on maximizing modularity, a metric for gauging the strength of the network's community structure. The algorithm is fast and scalable, making it an attractive option for analyzing large-scale networks. However, despite being quick and scalable, the Louvain algorithm may not perform well in detecting communities with non-overlapping nodes.

On the other hand, the label propagation algorithm (LPA), was suggested by Raghavan et al. [13] and considered a fast and scalable method for identifying communities in large-scale systems. A community is formed out of nodes that share the same label, by assigning a label to each node, and then iteratively propagating labels to its neighboring nodes. The LPA is effective at identifying communities with overlapping nodes since nodes might belong to several communities depending on the labels they receive. [7] Among the benefits of the LPA are simplicity and quickness. The algorithm only requires a single pass over the network, making it highly efficient for large-scale networks. However, the LPA can be sensitive to the initial labeling of nodes, and it may not perform well in networks with highly connected hubs.

The Girvan-Newman algorithm [14] is another popular approach for community identification. The betweenness centrality of an edge, which is a metric for how many shortest paths cross through it, is removed iteratively by the procedure. The procedure continues until the network is split into its component communities. Although the Girvan-Newman approach has been found to be effective at detecting communities in dynamic networks, it is computationally expensive and may not be suitable for studying large-scale networks. One disadvantage of the Girvan-Newman algorithm is that it may produce many small communities in a network.

In recent years, various methods have been proposed for identifying communities in dynamic networks that incorporate temporal information. These methods have applications in a wide variety of networks, including biological networks, social networks, and transportation networks, where the network structure changes over time, and communities evolve in response to external factors. For example, the Infomap [15] algorithm proposed by Rosvall and Bergstrom uses information theory (Random walks) to identify communities in dynamic networks. Nodes that exchange more information are more likely to be part of the same community, according to the algorithm which accomplishes this by running random walks on the network and then creating communities out of nodes that were visited by similar random walks. Infomap is particularly suited for examining changing networks since it considers the temporal ordering of edges to discover communities that are stable over time.

The Dynamic Stochastic Block Model proposed by Peel et al. [16] is another algorithm that models the development of communities over time using a probabilistic framework. The DSBM assumes that nodes are partitioned into blocks, and the probability of nodes having edges between them depends on whether or not they are members of that block. However, The DSBM extends the stochastic block model to include temporal dynamics, in contrast to the static form of the model. In particular, the DSBM assumes that the block memberships of nodes can change over time, and the probabilities of edges between nodes depend on both their current block memberships and their past block memberships. Therefore, it is helpful for investigating dynamic networks with non-stationary community structures because it can identify communities that change over time. [3]

In summary, the identification of communities in dynamic networks remains a complex and evolving research area that has garnered considerable interest among scholars. The choice of an optimal algorithm depends on several network characteristics, such as the characteristics of edges that connect the nodes, the size and density of the network, also the frequency and duration of network changes. Additionally, the choice of an algorithm for community detection may be impacted by the missing data, presence of noise, or outliers. Therefore, researchers should carefully consider these network characteristics when choosing an algorithm to detect communities in dynamic networks. Furthermore, scientists seeking to explore community identification in dynamic networks can use the algorithms discussed in this section as a solid starting point. [11]

3 Community Identification

Detection communities in a continuously evolving network are the process of identifying densely connected groups of nodes within a network that changes over time. Among the factors that make discovering communities challenging is the network's structure. For instance, there is a high level of local clustering in the ensemble of networks where the average distance between nodes increases at a gradual pace, indicating that nodes are more likely to be connected to their immediate neighbors. This property can make small-world networks more resilient to node failures or attacks. On the other hand, networks whose degree distribution follows a power law pattern, have a few highly connected nodes that are essential in preserving the network's structure and function. [2] Furthermore, certain topologies may be more adaptable to gradual changes, while others may be better suited to sudden changes or shocks. Additionally, changes in the network can happen in several ways, such as nodes joining or leaving the network, the creation or deletion of edges between nodes, or changes in the weights or strengths of existing edges. Thus, to detect communities in dynamic networks, researchers take into consideration of both the changing nature of the network and the structure of its communities. [8]

3.1 Community Detection Features

Community detection in dynamic networks has several key features that set apart it from community detection in static networks:

- Temporal aspect: Dynamic networks evolve over time, which means that community detection methods must take into account adjustments to the network's topology, node attributes, and edge weights.
- Community evolution: In dynamic networks, communities may expand, shrink, merge, split, or disappear over time. These changes should be tracked by community detection methods as well as provide insights into the community's life cycle.
- Time-aware algorithms: Community detection methods for dynamic networks should incorporate temporal information, either by analyzing the network at different time snapshots (evolutionary methods) or by direct accounting of time in the detection process (temporal methods).
- Scalability: Dynamic networks can be large and complex, along with the probability of nodes and edges that changing over time within the network. For that reason, community detection methods should be scalable to handle such networks efficiently.
- Robustness: While noise, missing data, and uncertainties in the dynamic network might impair the quality of discovering communities, community identification algorithms should be resistant to these factors.
- Evaluation and validation: Assessing the quality and stability of detected communities in dynamic networks requires appropriate evaluation metrics that consider both the network structure and temporal information.
- Interpretation and visualization: Understanding the detected communities and their evolution over time requires effective visualization techniques and tools that can help researchers gain insights into the network's structure, function, and dynamics.
- Adaptability: Community detection methods should be adaptable to different topologies of dynamic networks, such as social, biological, or technological networks, each with its unique characteristics and challenges.

3.2 Community Detection Categories

In network analysis, community detection is considered an effective tool to detect clusters of nodes that are more closely connected to one other than to other nodes in a network. These clusters, or communities, can reveal hidden structures and patterns within complex systems.[4] Community detection methods can be categorized into two types: Agglomerative Methods and Divisive Methods.

- Agglomerative methods involve adding edges to a graph one at a time, starting with the strongest edge and moving towards weaker edges. This procedure is repeated until the network is fully connected.
- Divisive methods work in the opposite direction, starting with a fully connected graph and removing edges one at a time until the network is fully separated into clusters.

Both methods have their advantages and disadvantages depending on the network's structure. [17]

3.3 Overlapping Communities

Overlapping communities in a network are a set of nodes that belong to multiple communities. Overlapping communities allow nodes to simultaneously belong to many communities, in contrast to conventional non-overlapping communities, where each node only belongs to one community. In real-world overlapping communities can be found in a wide variety of systems, ranging from social, biological, and transportation networks. In a social network, for example, overlapping communities may represent groups of individuals who share multiple interests or belong to multiple social circles. Finding communities that overlap is a challenging problem, that requires the development of specialized algorithms that can identify nodes that are a part of several communities while maintaining the integrity and coherence of each community. [18]

4 Community Detection Algorithms And Evaluation Measures

4.1 Louvain Algorithm

Definition 1 (Louvain). *The Louvain method, introduced by Blondel et al, is a community detection algorithm that uses agglomeration and hierarchical optimization. It employs a vertex mover (VM) procedure to enhance the modularity of the network at each level. The Louvain algorithm proceeds in two steps:*

- *In the first step, assign each node to its community to optimize modularity for small local communities. Then, iteratively merges communities to maximize the modularity score.*
- *In the second step, the communities discovered in the first step are considered individual nodes, and the same process is repeated until the maximum modularity score is reached, and no further merges can increase it. [1]*

To determine modularity gain, the modularity score of the network is compared before and after moving a node to a different community 1. The modularity score is the measure of how closely nodes are connected to each other within a community, compared to what might be expected by chance 2. [9]

$$\triangle Q = Q_a fter - Q_b efore \tag{1}$$

$$Q = (1/2p) * \sum_{i} j(A_ij - k_i * k_j/2p) * \delta(c_n, c_m) \tag{2}$$

where:

- *Aij: is the weight of the edge connecting nodes i and j.*
- *ki and kj: are the expected probabilities of edges connected to nodes i and j, respectively.*
- *p: is the sum of all edge weights in the network.*
- *cn and cm: are the communities to which nodes i and j belong, respectively.*
- *$delta(cn, cm)$ is a function that returns 1 if $ci = cj$ and 0 otherwise.*

is able to analyze complex systems since it can handle large-scale networks with billions of nodes. Furthermore, the algorithm is highly parallelizable, allowing it to take advantage of multi-core processors and distributed computing environments, further increasing its efficiency. The Louvain method has been demonstrated to be useful in identifying communities that are stable over time in dynamic networks, where the network topology changes over time. However, the algorithm may struggle to identify communities that are extremely volatile or that change rapidly over short periods of time. In such cases, other community detection algorithms may be more appropriate.

Despite its strengths, the Louvain algorithm has some limitations. One potential issue for the algorithm can get stuck in local optima, which can lead to suboptimal community assignments. To overcome this issue, researchers have developed a number of extensions and modifications to the algorithm, such as the Louvain-igraph implementation, which incorporates stochasticity to improve the algorithm's ability to escape local optima. [19]

4.2 Label Propagation Algorithm

Definition 2 (LP Algorithm). *The label Propagation algorithm is a semi-supervised machine-learning method that aims to predict the labels of unlabeled vertices of a graph through the distribution of the labels of labeled vertices. The algorithm operates under the presumption that vertices connected by edges in a graph have similar labels, and depending on the labels of its nearby vertices, it iteratively updates each vertex's label. At each procedure, each vertex is assigned the label that is most common among its neighbors, weighted by the strength of the edges that connect them. The algorithm continues to propagate the labels until a convergence criterion is reached. Label Propagation is a powerful and scalable algorithm that can perform various tasks, such as link prediction, social network analysis, and community detection. [13]*

The label Propagation algorithm is an efficient method for detecting societies in dynamic networks. It can analyze large datasets rapidly and accurately thanks to its scalability, which is its key feature. Label Propagation is a flexible and adaptable method for examining complex network structures because, in contrast, to many other community discovery algorithms, it doesn't necessitate prior knowledge regarding the sizes or the numbers of communities.

The label Propagation algorithm is an efficient method for detecting societies in dynamic networks. It can analyze large datasets rapidly and accurately thanks to its scalability, which is its key feature. Label Propagation is a flexible and adaptable method for examining complex network structures because, in contrast, to many other community discovery algorithms, it doesn't necessitate prior knowledge regarding the sizes or the numbers of communities.

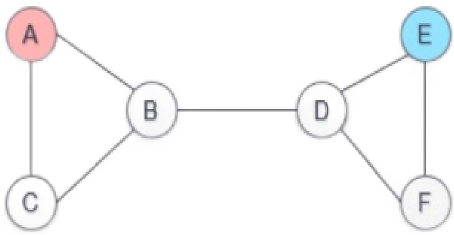

Fig. 1 Original Graph.

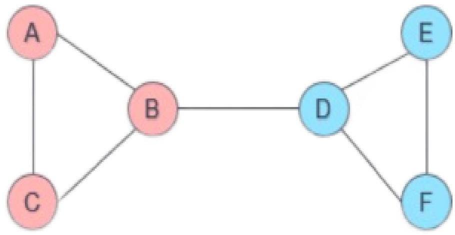

Fig. 2 Graph Applying LPA Algorithm.

Furthermore, it is typically regarded as a reliable and robust approach for detecting communities with a minimum of ambiguity and unpredictability. The algorithm ensures that each data point is assigned the label that is most frequently used by its neighbors, reducing the likelihood of random or spurious assignments. Additionally, the algorithm's convergence criterion and a maximum number of iterations provide a degree of control over the final community structure, helping to avoid overfitting or underfitting the data.

Nonetheless, Label Propagation has its limitations, just like any other algorithm. Particularly in highly dynamic networks where the community structure is continually changing, it may be difficult to recognize small or weakly connected groups. However, when the global network topology is complicated or poorly understood, the algorithm's reliance on local knowledge may produce inaccurate results. [20]

Table 1 Exploring the average modularity of Label Propagation algorithm using various values of *c* in different networks

Network	c = 1.0	c = 0.9	c = 0.8	c = 0.7	c = 0.6	c = 0.5	c = 0.4	c = 0
Polbooks	0.521	0.521	0.521	0.521	0.521	0.521	0.521	0.487
Netscience	0.804	0.804	0.804	0.804	0.804	0.804	0.804	0.798
Email	0.490	0.480	0.475	0.470	0.465	0.460	0.450	0.230
Eva	0.919	0.918	0.916	0.913	0.910	0.905	0.900	0.890
CA-GrQc	0.756	0.755	0.754	0.753	0.752	0.751	0.750	0.752
PGP	0.830	0.828	0.826	0.824	0.822	0.820	0.818	0.802
Facebook	0.854	0.853	0.853	0.852	0.852	0.851	0.850	0.845
Power Grid	0.101	0.100	0.100	0.099	0.098	0.097	0.096	0.086

In the table 1, each row represents a different network, while the columns correspond to various values of c. The values in each cell represent the average modularity achieved by LPA on that network with the given value of c. The last column shows the as a baseline for comparison. The networks included in this table are just examples and were chosen to demonstrate the varying performance of LPA on different networks and with different values of c.

4.3 Girvan-Newman Algorithm

Definition 3 (Girvan-Newman). *The Girvan-Newman algorithm is a hierarchical divisive method for detecting communities in complex systems. The algorithm works by: [14]*

- *Compute the betweenness centrality for every edge in the network.*
- *Eliminate the edges with the greatest betweenness centrality from the network.*
- *Recalculate the betweenness centrality for every edge in the remaining network.*
- *Repeat steps two and three until a termination condition is satisfied. This criterion could be a desired number of communities, a desired community size, or when the network is completely disconnected.*
- *Divide the network into communities by using the connected components that persist after the edges are eliminated.*

The Girvan-Newman algorithm has been widely used to determine the underlying structures of various types of networks. Along with Its ability to detect communities in a continuously changing network is one of the features that make required. To achieve this, the edges with the highest interfacial centrality are progressively removed, thus revealing the modular structure of the network. 1

The Girvan-Newman approach also has the advantage of producing high-quality community detection results while avoiding unpredictability. This is due to the fact that it uses a bottom-up approach, where smaller communities are merged to form larger ones, based on the network's underlying structure. This process is guided by the modularity measure, which is used to assess the quality of the communities at each step. [10]

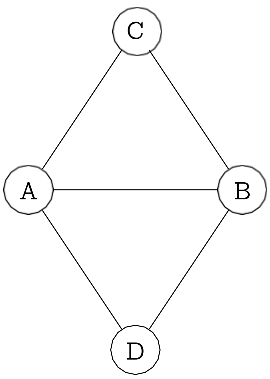

Fig. 3 Original Graph.

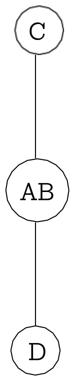

Fig. 4 Graph After Applying The Girvan-Newman method.

However, The Girvan-Newman algorithm can have several drawbacks, such as the computational complexity and the sensitivity to the choice of the threshold for edge removal. 2 Furthermore, it may struggle with networks that have overlapping communities, or those that exhibit strong community structures at different scales. [11]

4.4 Infomap Algorithm

Definition 4 (Infomap). *The concept behind the Infomap algorithm involves the optimization of a quantity known as the "map equation," which aims to assess the excellence of a given community partition within a network. At its core, the algorithm works by iteratively compressing the network into increasingly smaller modules, while minimizing the information loss between nodes within the same module. Specifically, by constructing a hierarchical tree structure that represents the network's modular organization. The algorithm endeavors to locate the partition that diminishes the map equation, with every tier of the tree corresponding to a distinct network community partition. [15]*

The map equation is a mathematical formula that describes the balance between two competing factors in community detection : the quality of the individual modules (i.e., how tightly connected the nodes within a module are) and the information cost of encoding the module structure (i.e., how much information is required to represent the community partition).

$$S(M) = q \cdot H(Q) + \sum_{i=1}^{n} p_i \cdot H(P_i|Q) \quad (3)$$

- *S(M) is the anticipated code length for the random walk over the network.*
- *q is the likelihood of transitioning from one community to another.*
- *H(Q) is the entropy of the community distribution Q, which is defined as*

$$\sum_{c \in C} q_c \cdot \log_2(q_c) \quad (4)$$

where:

- *C: represents the set of all network communities and*
- *qc: represents the likelihood of ending up in community c.*
- *m: is the count of nodes existing in the network.*
- *pi: represents the likelihood of initiating a random walk from node i in the network.*
- *H(Pi/Q): refers to the amount of uncertainty, on average, about which community a node i belongs to given the distribution of communities Q. It is the conditional entropy of node i with respect to the community distribution Q, which is defined as*

$$\sum_{c\in C} q_c \cdot p_{i,c} \cdot \log_2(p_{i,c}) \tag{5}$$

To find the optimal partition, the Infomap algorithm first assigns each node to its own community 3 and subsequently merges communities based on the improvement in the map equation 3. This process continues until no further improvements can be made. 4

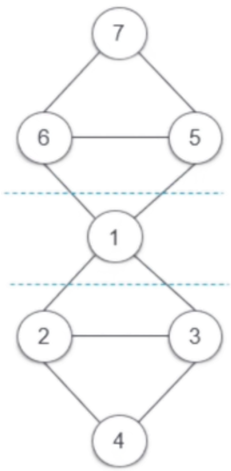

Fig. 5 Original Graph.

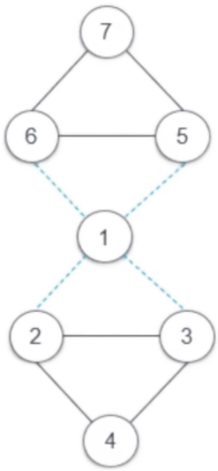

Fig. 6 Applying Infomap Algorithm.

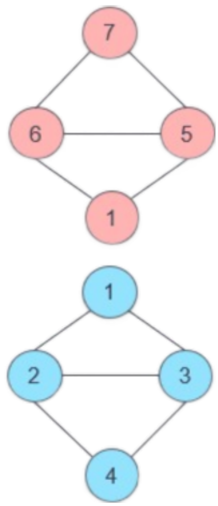

Fig. 7 Graph after Infomap Algorithm.

The Infomap algorithm is a widely used community detection algorithm that has been applied to a variety of complex networks. The fundamental concept behind the as a message or code and minimizing the anticipated length of this code to find the most favorable division of a network into distinct communities.

One of the main advantages of the Infomap algorithm is its ability to detect communities in large and complex networks. It has been demonstrated to outperform in superior accuracy and computing efficiency over several existing community detection techniques. Moreover, the hierarchical nature of the algorithm allows it to identify communities at different levels of granularity, which is useful in comprehending the organization of complex networks.

Nonetheless, one restriction of the algorithm is its dependence on the initial conditions, which can lead to variable outcomes based on the initial position of the random walker. Moreover, the algorithm may not be suitable for dynamic networks where the network configuration changes frequently over time. In dynamic networks, the community structure may change rapidly over time, and the Infomap algorithm may not be able to keep up with these changes. [21]

4.5 The Dynamic Stochastic Block Model Algorithm

Definition 5 (DSBM). *The Dynamic Stochastic Block Model algorithm is a computational method used to identify communities or clusters in a dynamic network. It works by assuming that nodes in a network 5 are divided into latent groups or blocks, which can change over time. The DSBM algorithm uses statistical models to infer the underlying blocks and their evolution over time based on the observed network structure. In other words, it seeks to discover the groups of nodes that are more likely to interact with each other, and how these groups change over time. The algorithm is labeled as "dynamic" because it considers the temporal dimension of the network, and "stochastic" because it relies on probability models to describe the network's structure. [16]*

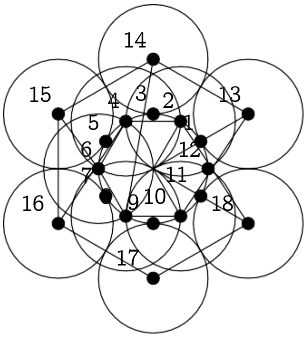

Fig. 8 The original network before applying The Dynamic Stochastic Block Model algorithm.
The Dynamic Stochastic Block Model (DSBM) algorithm is a generative model that assumes that nodes in the network are assigned to different communities and that

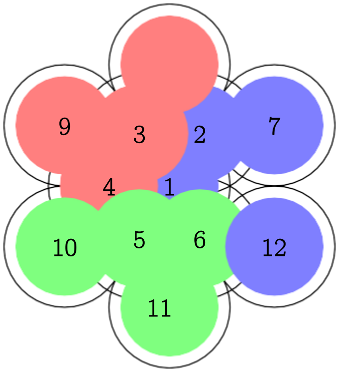

Fig. 9 The network after applying The Dynamic Stochastic Block Model algorithm.

the connections between nodes are probabilistically determined by the community assignments 6.

The DSBM algorithm has a significant advantage in that it can represent the network's progression over time, enabling the identification of alterations in the community configuration as the network evolves. This is particularly useful in dynamic networks, where the relationships between nodes may change over time, and communities may form and dissolve. Furthermore, it is also highly efficient, making use of Bayesian inference techniques to estimate the community assignments and network parameters. This allows for fast and accurate community detection even in large and complex networks.

However, as with any algorithm, there are limitations to the DSBM. One potential issue is the difficulty in choosing the appropriate quantity of communities to model. If the quantity of communities is set too high or too low, it may lead to inaccurate results. Another potential limitation is the assumption of a static community structure within each time interval, which may not always hold in dynamic networks. In terms of avoiding randomness in words and minimizing perplexity, it is important to note that the DSBM is a mathematical model and not a language model. Therefore, it does not involve the use of words or natural language processing. Nevertheless, in the context of identifying communities in continuously changing networks, it is important to ensure that the model is appropriate for the data being analyzed and that the results are interpreted in a meaningful way. [23]

5 Significant Findings in the Algorithms for Community Detection

The table 2 provides a comparison of community detection algorithms, highlighting important observations regarding their limitations, complexity, improvements, and features related to modularity optimization and overlapping communities.

The Louvain algorithm is a fast algorithm that can handle large datasets with a complexity of O(N log N). However, it has a resolution limit, which means it cannot

Table 2 Comparison of community detection algorithms

Algorithm	Limitations	Complexity	Improvement	Modularity Optimization	Overlapping Communities
Louvain	Resolution limit, sensitivity to initialization	O(N log N)	Louvain-igraph algorithm, resolution limit correction	Yes	Yes
Label Propagation	Sensitivity to initialization, resolution limit	O(N)	Asynchronous label propagation, spectral initialization	Yes	Yes
Girvan-Newman	High computational cost, resolution limit, sensitivity to community size	$O(N_3)$	Fast algorithms based on edge-betweenness, resolution limit correction	Yes	No
Infomap	Sensitivity to resolution limit, overfitting	O(N log N)	Regularization, bias correction	Yes	Yes
Dynamic Stochastic Block Model	Computationally intensive, sensitivity intensive, sensitivity	$O(T\ N_2)$	Variational Bayes inference, model selection	Yes	Yes

detect communities that are smaller than a certain size. It is also sensitive to initialization, and its performance can be improved by using Louvain-igraph algorithm and resolution limit correction. The Louvain algorithm is capable of detecting overlapping communities.

Label Propagation algorithm has a lower computational cost than Louvain, with a complexity of O(N). However, it is sensitive to initialization and has a resolution limit. The algorithm can be improved by using asynchronous label propagation and spectral initialization. The Label Propagation algorithm is capable of detecting overlapping communities.

The Girvan-Newman algorithm is computationally intensive, with a complexity of $O(N^3)$. It also has a resolution limit and is sensitive to community size. Fast algorithms based on edge-betweenness can improve their performance, but they cannot detect overlapping communities.

The Infomap algorithm can manage a resolution limit and has a computational complexity of O(N log N). However, it is sensitive to overfitting and resolution limits. Regularization and bias correction can be used to improve its performance. The Infomap algorithm is capable of detecting overlapping communities.

The Dynamic Stochastic Block Model algorithm is computationally intensive, with a complexity of $O(T\ N^2)$, where T is the number of time steps. It is sensitive to model assumptions and requires variational Bayes inference and model selection for improved performance. The Dynamic Stochastic Block Model algorithm can detect overlapping communities. [22]

6 conclusion

Community detection algorithms play a crucial role in identifying hidden structures and relationships within complex networks. The comparison table 2 highlights popular community detection algorithms such as Louvain, Label Propagation, Girvan-Newman, Infomap, and Dynamic Stochastic Block Model. Each algorithm has its strengths and weaknesses, and selecting the appropriate algorithm depends on the dataset's characteristics and research questions. Louvain and Label Propagation algorithms are suitable for large datasets and can detect overlapping communities. The Girvan-Newman algorithm can detect non-overlapping communities but is computationally intensive. The Infomap algorithm can handle resolution limits and detect overlapping communities but is sensitive to overfitting. The Dynamic Stochastic Block Model algorithm is suitable for dynamic networks but is computationally intensive and requires model selection.

In conclusion, community detection algorithms present a potent approach to analyzing intricate networks, also selecting the appropriate algorithm requires careful consideration of the dataset's characteristics and research objectives.

References

[1] ZhaoYang, Ren´eAlgesheimer Claudio J.Tessone. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks, 2016.

[2] RÉMY CAZABET GIULIO ROSSETTI. (2018). Community Discovery in Dynamic Networks: A Survey.

[3] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong Rong Jin. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach.

[4] Santo Fortunato. (2010). Community detection in graphs.

[5] M. E. J. Newman. (2006). Modularity and community structure in networks.

[6] Andrea Lancichinetti, Filippo Radicchi, Jose´ J. Ramasco, Santo Fortunato. (2011). Finding Statistically Significant Communities in Networks.

[7] Jierui Xie, Mingming Chen, Boleslaw K. Szymanski . (2013). LabelRankT: Incremental Community Detection in Dynamic Networks via Label Propagation.

[8] Mrs. Kulkarni Varsha, Dr. Kiran Kumari Patil. (2020). An Overview of Community Detection Algorithms in Social Networks.

[9] Ernesto Estrada. (2011). Community detection based on network communicability.

[10] Ljiljana Despalatovic, Tanja Vojkovic, Damir Vukicevi. (2014). Community structure in networks: Girvan-Newman algorithm improvement.

Detection: Comparison of State of the Art Algorithms.

[12] Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

[13] Raghavan, U. N., Albert, R., Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. Physical review E, 76(3), 036106.

[14] Girvan, M., Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12), 7821-7826.

[15] Rosvall, M., Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. Proceedings of the National academy of Sciences, 105(4), 1118-1123.

[16] Peel, L., Delvenne, J. C., Lambiotte, R. (2017). The multilayer temporal block model reveals the dynamic structure of HIV infection among MSM. Journal of The Royal Society Interface, 14(131), 20170220.

[17] Majid Arasteh · Somayeh Alizadeh (2018). A fast divisive community detection algorithm based on edge degree betweenness centrality.

[18] Jaewon Yang, Jure Leskovec, (2013). Overlapping Community Detection at Scale: A Nonnegative Matrix Factorization Approach.

[19] Jicun Zhang , Jiyou Fei , Xueping Song , and Jiawei Feng , (2021), An Improved Louvain Algorithm for Community Detection.

[20] Jierui Xie and Boleslaw K. Szymanski. (2011). Community Detection Using A Neighborhood Strength Driven Label Propagation Algorithm.

[21] Cuijuan Wang, Wenzhong Tang, Bo Sun, Jing Fang , Yanyang Wang . (2015) Review on Community Detection Algorithms in Social Networks.

[22] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborova . (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications.

[23] Mrs. Kulkarni Varsha , Dr. Kiran Kumari Patil , An Overview of Community Detection Algorithms in Social Networks