Chapter 6

Conclusion and Future Research Direction

6.1 Conclusion

The prevalence of networks extends throughout nature, encompassing both living organisms and inanimate entities. Its ubiquity presents ample opportunities to glean meaningful properties and characteristics from it. The identification of groups, commonly referred to as communities, within a network has garnered considerable research attention. A community comprises a collection of nodes that share common behavioral attributes, thus furnishing valuable structural and behavioral insights into any network. Such communities are structured to exhibit a high ratio of internal to external connections.

Traditionally, a network is represented as a graph consisting of nodes interconnected by edges. However, this definition necessitates modification to incorporate another crucial dimension: time. The temporal aspect enables a more nuanced understanding of networks by incorporating valuable temporal information. Social networks, pervasive in the real world, exemplify this concept. Instances of such social networks include collaborative research groups, corporate communication networks, online friendship circles, behavior-based animal networks, among others. They can be conceptualized as graphs, with nodes representing network entities and edges symbolizing the relationships between them. Due to their widespread real-world applicability, social networks have attracted significant research interest.

The size of the global datasphere has experienced exponential growth, expanding from 0.5 exabytes in 2009 to 18 exabytes in 2018. This exponential increase, driven by the proliferation of internet-based services, has resulted in larger and more dynamic social networks. These networks not only possess vast volumes of data but also exhibit temporal evolution. The continuous influx of data gives rise to dynamic networks that evolve over time. Nowadays, most networks are dynamic in nature and lack a fixed topology. As such, evolving networks introduce new considerations even as efforts to extract information from their previous states are underway. Storing or processing such networks in their entirety at once is impractical; instead, an efficient approach necessitates the sampling of data into smaller subsets.

The thesis aims to introduce novel techniques for identifying communities within dynamic social networks. It comprises three algorithmic approaches aimed at addressing the objective functions within a mathematical framework that incorporates the dynamic nature of communities across diverse evaluation criteria. The first approach employs a tree-based methodology leveraging social network properties to inform community decision-making. The second approach utilizes multi-objective optimization techniques, drawing on knowledge of node neighborhoods in community decision-making processes. The final approach focuses on designing an algorithm to delineate the fuzzy membership of a node within a community. These proposed algorithmic approaches offer efficiency in identifying communities and monitoring their evolution over time.

6.1.1 Key Research Findings

Chapter 3 addresses the community detection problem in dynamic social networks using a tree-based algorithm (TCD2) that leverages connectedness and influence. The main contributions of the work are:

- **Utilization of Network Properties:** The algorithm identifies closely related clusters by utilizing the connectedness and influence properties in a dynamic setting. It employs a tree-like structure to incorporate these properties, with an exemplary graph provided to illustrate its functionality and complexity analysis.

- **Experimental Validation:** The proposed algorithm is validated through experiments on both real-world and synthetic social networks, using several well-known evaluation metrics. Comparative analysis is conducted against four state-of-the-art static and dynamic algorithms. The performance of TCD2 is analyzed under static and dynamic conditions, with considerations of level counts, window size, and mixing parameters.

- **Performance and Flexibility:** Results show that TCD2 is more effective in identifying true community groups compared to state-of-the-art algorithms. It offers flexibility in managing community quality and accuracy without requiring a predetermined number of communities. TCD2 provides a balance between quality and accuracy in community detection.

Chapter 4 focuses on community detection in dynamic networks, addressing the limitations of traditional methods that rely on snapshots of the network. Instead, it treats the network as a stream of events, which eliminates the need for mapping and minimizing temporal shifts between consecutive snapshots. This approach allows for real-time operations on each event and its involved nodes, significantly reducing complexity and improving algorithm speed. Key contributions of the work include:

- **Objective Functions Inspired by Network Properties:** The proposed work utilizes three objective functions inspired by inherent network properties. These functions are optimized using the Pareto front within a relatively smaller search space, rather than the entire network.

- **Temporal Network Model Implementation:** The algorithm operates within a temporal network model, allowing for real-time community detection at each event without waiting for the completion of a timestamp.

- **Evaluation on Diverse Datasets:** The performance of the algorithm is evaluated on both connected and disconnected datasets, including real and synthetic data, ensuring a comprehensive analysis of its effectiveness across different network types.

- **Comparative Analysis with State-of-the-Art Algorithms:** A comparative analysis is presented, showing how the proposed algorithm measures up against various state-of-the-art algorithms from the literature.
- **Superior Performance in Accuracy Metrics:** The assessment indicates that the algorithm performs better than most state-of-the-art methods, particularly excelling in accuracy metrics. While quality metrics for synthetic datasets show slight lags, the algorithm still maintains competitive performance.

Chapter 5 introduces a community detection algorithm tailored for dynamic networks, leveraging a fuzzy membership function to determine the extent of a node's association with any community. The main contributions include:

- **Novel Fuzzy Membership Function:** A novel fuzzy membership function is introduced, considering the influence of a node's neighborhood and the neighborhood of its neighbors. These factors provide sufficient information to analyze community transitions over time.
- **Dynamic Social Network Analysis:** The algorithm addresses the community detection problem in dynamic social networks, presenting the global community structure at time t. Upon the arrival of a new node, it is assigned to a community based on the membership function's value.
- **Extensive Evaluation:** The proposed algorithm is evaluated using seven well-known quality and accuracy metrics across three real and five synthetic datasets.
- **Comparative Analysis:** A comparative analysis is conducted against ten state-of-the-art algorithms, demonstrating that the proposed algorithm outperforms the others in most metrics.

6.2 Future Directions

As the conclusions of this thesis are drawn, it becomes evident that further investigation is warranted, with numerous promising directions for future research emerging from this work. Areas that merit further exploration and investigation are as follows:

Temporal information: Techniques proposed in this thesis follows temporal network model which considers network as an evolving entity. More complex dynamics may appear in real world networks such as removal of an edge/ node or admission of an edge for a fixed period of time. These scenario increases the complexity of network model and also imposes challenges to algorithms.

Overlapping communities: Study presented here focuses on disjoint communities. However, in real world networks single node may belong to multiple communities. Chapter 5 puts a notion of membership-based community identification but it considers non-overlapping communities only which can further be extended for overlapping communities.

Improved performance: Further improvement in performance of algorithms is a well-known ambition of researchers. As the dynamic networks require light and fast computation, parallel computation or dynamic programing can be some prudent options for this purpose. These computations may also allow researchers to study networks with millions of nodes and edges.