Advancements in Community Detection and Analysis Methods: A Comprehensive Review of Networks of Overlapable Communities

Bashar Mohammed Tuama[1]
[1]Computer Science, Altinbas University, Turkey Istanbul, Ministry of Health / Salah Al-Din Health Department / Balad Sector. Email: bashar.balad@gmail.com

Laith Ali Abdulsahib[2]
[2]Department of Computer Science/ Faculty of Education for Women/ University of Kufa/ Iraq Email: laith.albaldawi@uokufa.edu.iq

ABSTRACT

Community detection is a fundamental problem in complex network analysis, aiming to uncover cohesive groups of nodes with dense internal connections and sparse connections to nodes outside the group. While traditional community detection methods focus on identifying disjoint communities, many real-world networks exhibit overlapping community structures, where nodes can belong to multiple communities simultaneously. Overlapping communities are prevalent in diverse systems, such as social networks, biological networks, and technological networks, where entities often participate in multiple contexts or functionalities.

This survey comprehensively reviews the advancements in overlapping community detection methods. We explore various community detection techniques, including modularity-based algorithms, hierarchical clustering approaches, and probabilistic models, each with unique capabilities in identifying overlapping communities. Moreover, we delve into node membership-based approaches, link clustering methods, and fuzzy clustering algorithms, explicitly addressing nodes' participation in multiple communities. Evaluating the performance of overlapping community detection methods is essential for understanding their effectiveness. We discuss various evaluation metrics commonly used in the literature and assess their suitability for measuring the quality of overlapping community partitions.

Real-world applications of overlapping community detection in diverse domains are also presented. These applications demonstrate the significance of uncovering overlapping structures in complex systems, leading to the identification of functional modules and hidden patterns in real-world networks.

This survey offers valuable insights for researchers and practitioners in complex network analysis by providing a systematic and in-depth analysis of overlapping community detection techniques. Understanding overlapping communities is essential for gaining deeper insights into complex systems' modular organization and functionalities, enabling researchers to make informed decisions and discoveries across various domains.

Keywords: Community detection, Complex networks, Overlapping communities, Network analysis, Algorithm evaluation

1. Introduction

Complex networks, ranging from social systems to biological and technological structures, have become ubiquitous in various domains (Newman, 2010; Albert & Barabási, 2002; Barabási & Oltvai, 2004). The study of these networks has provided valuable insights into the underlying organization and dynamics of interconnected entities. One key aspect of understanding complex networks is the identification of communities or clusters, which are groups of nodes with dense connections among themselves and sparse connections to nodes outside the group (Fortunato, 2010; Girvan & Newman, 2002). Community detection is a fundamental problem in network analysis that has attracted significant research attention over the past few decades.

Community detection algorithms have traditionally focused on identifying disjoint communities, where each node belongs exclusively to one community (Fortunato, 2010). However, many real-world networks exhibit a more intricate organization, with nodes participating in multiple communities simultaneously. This phenomenon gives rise to networks with overlappable communities, where nodes can have fractional memberships in more than one community (Palla et al., 2005). Overlapping communities are particularly prevalent in social networks, where individuals often belong to multiple social circles, and in biological networks, where genes or proteins may be involved in multiple biological processes (Palla et al., 2007; Lancichinetti et al., 2009).

Detecting and characterizing overlapping communities is crucial for understanding complex systems' modular structure and functionalities (Fortunato, 2016; Lancichinetti et al., 2009). Overlapping community detection methods offer a more nuanced representation of node-community relationships, providing a deeper understanding of how entities interact and function in complex networks. By uncovering overlapping communities, researchers can reveal hidden patterns of connectivity, identify functional modules, and explore the diverse roles nodes play within different contexts (Ravasz & Barabási, 2003; Blondel et al., 2008).

The survey explores community detection techniques, including modularity-based algorithms, hierarchical clustering, and probabilistic models, for identifying overlapping communities. Noteworthy examples include the Louvain algorithm and Infomap. Section 2 delves into overlapping community detection methods, such as OSLOM, COPRA, LinkComm, and fuzzy clustering. Evaluation metrics like Modularity, NMI, and VI are discussed in Section 3 to assess method performance. Section 4 highlights real-world applications in diverse domains. The survey provides valuable insights for researchers, offering a comprehensive understanding of overlapping communities and their significance in complex networks. It serves as a valuable resource, unlocking new opportunities for research and application in the field of complex network analysis.

This survey aims to provide a comprehensive review of the advancements in overlapping community detection methods. By exploring state-of-the-art techniques, algorithms, evaluation metrics, and applications, this survey contributes to the advancement of research in complex networks. The survey is organized into sections, each focusing on different aspects of overlapping community detection, providing a systematic and in-depth analysis of the field.

1. Community Detection Techniques:

Community detection techniques are fundamental tools for understanding the structural organization of complex networks. These methods aim to identify cohesive groups of nodes within a network, known as communities or clusters, based on their patterns of connectivity. Traditional community detection methods have primarily focused on identifying disjoint communities, where each node belongs exclusively to one community. However, real-world networks often exhibit a more complex structure, with nodes participating in multiple communities simultaneously. This phenomenon gives rise to networks with overlappable communities, where nodes can have fractional memberships in more than one community. The presence of overlapping communities is particularly prevalent in social networks, where individuals may belong to multiple social circles, and in biological networks, where genes or proteins may be involved in multiple biological processes.

1.1 Modularity-Based Methods

Modularity optimization is one of the most widely used approaches for community detection (Newman, 2016). It involves optimizing a modularity function that quantifies the difference between the number of edges within communities and the expected number of edges in a random network. Several variations of modularity-based methods have been proposed, each offering different advantages for detecting overlapping communities.

The Louvain algorithm is a popular modularity-based method that efficiently partitions large networks into communities (Blondel et al., 2008). It uses a greedy optimization technique to merge nodes into communities to maximize the modularity score iteratively. The Louvain algorithm is known for its speed and scalability, making it suitable for real-world applications.

Another notable modularity-based method is the Infomap algorithm (Rosvall & Bergstrom, 2008). Instead of directly maximizing modularity, Infomap focuses on minimizing the description length of a random walker's trajectory on the network. This approach results in the identification of communities with overlapping membership, allowing for the detection of finer-grained structures in the network.

1.2 Hierarchical Clustering Approaches

Hierarchical clustering algorithms construct a tree-like structure of communities, enabling the identification of communities at different levels of granularity (Fortunato, 2016). These algorithms have been extended to handle overlapping communities, where nodes can belong to multiple communities at various levels of the hierarchy.

The COPRA (Community Overlap PRopagation Algorithm) algorithm is one such method that iteratively propagates community labels to neighbouring nodes, identifying overlapping structures (Gregory, 2010). COPRA considers both the community membership of a node and the memberships of its neighbors during the propagation process, resulting in accurate detection of overlapping communities.

The CPM (Clique Percolation Method) algorithm is another hierarchical clustering approach that identifies overlapping communities based on the presence of k-cliques as atomic building blocks (Palla et al., 2005). Overlapping communities are then formed by combining cliques that share common nodes.

1.3 Probabilistic Models

Probabilistic models provide a flexible framework for community detection, accommodating nodes with probabilistic membership to multiple communities. These models allow for a more nuanced representation of node-community relationships and have been successfully applied to networks with overlapping communities.

The Stochastic Block Model (SBM) is a widely used probabilistic model for community detection (Peixoto, 2017). SBM assumes that the network's edges are generated based on a set of latent community assignments, and nodes within the same community share similar connection probabilities. This model allows for identifying overlapping communities with varying degrees of interconnectedness.

The Latent Space Model is another probabilistic approach that represents nodes and communities as points in a latent space, where closer points indicate higher probabilities of node membership in corresponding communities (Karrer & Newman, 2011). The Latent Space Model offers a powerful way to model overlapping communities and has been successfully applied to various real-world networks.

2. Overlapping Community Detection.

Overlapping community detection techniques specifically address the simultaneous identification of nodes participating in multiple communities. These methods aim to assign fractional membership values to each node, reflecting their degree of participation in each community. By capturing the intricate relationships between nodes and communities, overlapping community detection methods provide a more detailed and accurate representation of the network's structure.

2.1 Node Membership-Based Approaches

Node membership-based approaches assign fractional membership values to each node, indicating their degree of participation in different communities. These methods have demonstrated effectiveness in capturing the complexity of overlapping communities.

The OSLOM (Order Statistics Local Optimization Method) algorithm employs statistical significance to identify communities with overlapping nodes (Lancichinetti et al., 2011). OSLOM starts with small communities and iteratively grows them by assessing the statistical significance of adding community nodes. This approach ensures the robustness of detected communities and is particularly suitable for networks with complex and hierarchical overlapping structures.

The COPRA algorithm, mentioned in Section 1, is another node membership-based approach that iteratively refines the community assignments of nodes (Gregory, 2010). By considering both a node's membership and the memberships of its neighbours, COPRA accurately detects overlapping communities.

2.2 Link Clustering Methods

Link clustering methods focus on identifying groups of links that exhibit similar connectivity patterns, leading to the discovery of overlapping communities (Ahn et al., 2010). These methods provide an alternative perspective on community detection and have successfully uncovered overlapping structures in networks.

The LinkComm algorithm employs random walks to uncover overlapping link communities in networks (Raghavan et al., 2007). By considering the structure of links rather than node attributes, LinkComm efficiently captures the intricate relationships between nodes and their interconnections.

2.3 Fuzzy Clustering Algorithms

Fuzzy clustering algorithms extend traditional community detection methods by assigning nodes to communities with fuzzy membership values (Xie et al., 2013). In fuzzy clustering, nodes can belong to multiple communities with varying degrees of membership, providing a more nuanced representation of node-community relationships.

The Fuzzy C-Means algorithm is a well-known method for fuzzy clustering (Bezdek et al., 1984). It partitions nodes into multiple communities based on minimizing the weighted sum of squared deviations from the assigned community centers. Fuzzy C-Means has been extended to handle overlapping communities, allowing for detecting nodes with multiple community memberships.

Fuzzy Modularity is another important algorithm in the fuzzy clustering category, which extends modularity optimization to accommodate fuzzy memberships (Lancichinetti & Fortunato, 2009). This allows for the identification of overlapping communities with varying degrees of interconnectedness.

3. Evaluation Metrics for Community Detection:

When it comes to assessing the performance of community detection algorithms, it's essential to evaluate the quality of the detected communities. This means utilizing a range of evaluation metrics to accurately measure the effectiveness of community detection methods in networks where communities may overlap. By doing so, we can determine the accuracy of these algorithms and ensure they are providing reliable results.

3.1 Normalized Mutual Information (NMI)

Normalized Mutual Information (NMI) is a widely used metric to assess the similarity between the detected communities and the ground truth partition (Strehl & Ghosh, 2002). NMI measures the mutual information between the true and detected community assignments and provides a normalized score between 0 and 1, with 1 indicating a perfect match between the detected and ground truth communities. However, NMI may not be suitable for evaluating algorithms detecting overlapping communities, as it assumes that each node can only belong to a single community.

3.2 Variation of Information (VoI)

The Variation of Information (VoI) metric quantifies the information gain or loss between the true and detected community assignments (Meilă, 2007). It is based on the concepts of entropy and mutual information and provides a comprehensive evaluation of the performance of community detection algorithms, especially in cases of overlapping communities. A lower VoI value indicates a higher agreement between the detected and true community assignments.

3.3 F1-Score

The F1-Score is a widely used metric in machine learning and is applicable to community detection evaluation as well (Yang & Leskovec, 2012). It combines precision and recall, providing a balance between false positives and false negatives. The F1-Score is particularly useful for evaluating algorithms in imbalanced datasets, where some communities may be significantly smaller or larger than others.

3.4 Overlapping NMI (ONMI)

Overlapping Normalized Mutual Information (ONMI) is an extension of NMI specifically designed to evaluate the detection of overlapping communities (Xie et al., 2013). ONMI compares the overlapping membership of nodes in the true and detected communities, providing a measure of how well the detected communities capture the actual overlap. ONMI takes into account the fractional memberships assigned to nodes, making it more suitable for evaluating algorithms that handle networks with overlappable communities.

4. **Applications of Overlapping Community Detection**.

The detection of overlapping communities has numerous applications in various domains, providing valuable insights into the organization and dynamics of complex systems.

4.1 Social Networks

Social networks are a prominent area where overlapping communities have significant applications. In social networks, individuals often belong to multiple social circles, such as family, friends, colleagues, and interest groups. Detecting overlapping communities can help identify influential individuals who bridge multiple communities and facilitate information flow (Barber & Clark, 2019). Additionally, overlapping community detection can aid in targeted marketing strategies by identifying individuals with diverse interests and preferences.

4.2 Biological Networks

Biological networks, such as protein-protein interaction networks and gene regulatory networks, are characterized by intricate and interconnected functional modules. Overlapping community detection in biological networks can reveal proteins or genes involved in multiple biological processes (Lancichinetti et al., 2012). This information is invaluable for understanding the cross-talk between different biological pathways and uncovering the underlying mechanisms governing complex biological systems.

4.3 Citation Networks

In citation networks, overlapping communities can provide insights into the interdisciplinarity of scientific research. Papers belonging to multiple research topics or span different fields can be identified as key drivers of innovation and knowledge transfer (Ravasz & Barabási, 2003). Overlapping community detection in citation networks can facilitate the identification of emerging interdisciplinary research areas and promote collaboration across different scientific domains.

4.4 Co-authorship Networks

Researchers often collaborate on multiple research projects in co-authorship networks, leading to networks with overlapping communities. Identifying these overlapping communities can help uncover the dynamics of research collaborations, identify prolific researchers who contribute to diverse fields, and assist in forming research teams for specific projects (Lancichinetti & Fortunato, 2009).

4.5 Recommender Systems

Recommender systems aim to provide personalized and relevant recommendations to users based on their preferences and interests. Overlapping community detection can enhance the performance of recommender systems by identifying users with diverse interests (Zhang et al., 2019). Recommending items from different overlapping communities can increase user satisfaction and encourage exploration of new content.

5. Challenges and Future Directions:

While the field of overlapping community detection has witnessed significant progress, several challenges and future directions warrant consideration.

5.1 Computational Complexity

Many advanced community detection algorithms designed for overlapping communities exhibit high computational complexity. As network sizes grow, the computational cost of these algorithms may become a limiting factor. Addressing this challenge requires the development of scalable algorithms capable of efficiently detecting overlapping communities in large-scale networks.

5.2 Community Resolution

The resolution of detected communities is another crucial challenge in community detection. Overlapping community detection methods must strike a balance between identifying fine-grained community structures and avoiding over-partitioning the network. Improving the resolution of detected communities while preserving their interpretability remains an active area of research.

5.3 Definition of Overlaps

The concept of overlapping communities needs a universal definition, and different algorithms may produce varying notions of overlaps. Clarifying and standardizing the definition of overlaps in the context of community detection is essential for accurately comparing and evaluating different algorithms' performance.

5.4 Integration of Domain-Specific Information

In many real-world applications, networks often contain domain-specific information, such as node attributes, temporal dynamics, or geographical location. Integrating such information into the community detection process can enhance the accuracy and relevance of detected communities. Future research should explore methods for incorporating domain-specific information in overlapping community detection algorithms.

5.5 Real-Time Analysis

The dynamics of complex networks are often subject to change over time. Overlapping community detection methods must be adaptable to dynamic networks and capable of providing real-time analysis and updates. Developing algorithms that can efficiently detect and track changes in overlapping communities will be essential for time-critical applications.

Conclusion:

Significant progress has been made in studying community detection in complex networks, with a particular focus on overlappable communities. This comprehensive survey has explored key concepts, algorithms, evaluation metrics, and real-world applications related to overlapping community detection. Advancements in this field have underscored the importance of understanding and revealing overlapping structures in various domains, such as social networks, biological networks, and citation networks.

While substantial strides have been taken, challenges and future directions have also been identified. Addressing computational complexity, community resolution, and the definition of overlaps will be critical for enhancing the efficiency and accuracy of overlapping community detection methods. Integrating domain-specific information and developing real-time analysis capabilities will further expand the applicability of these techniques to dynamic systems.

As the study of complex networks continues to evolve, further research in overlapping community detection is expected to provide valuable insights into the intricate organization and functionality of complex systems. This promising field holds the potential to unlock new knowledge and applications across disciplines, including social sciences, biology, and computer science. By offering a thorough review of key concepts and advancements, this survey contributes to the advancement of research in complex networks and aids in the pursuit of a deeper understanding of overlappable communities.

References:

1. Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. Nature, 466(7307), 761-764. https://doi.org/10.1038/nature09182
2. Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal

multiscale complexity in networks. Nature, 466(7307), 761-764.
3. Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 47-97.
4. Ball, B., Karrer, B., & Newman, M. E. (2011). Efficient and principled method for detecting communities in networks. Physical Review E, 84(3), 036103. https://doi.org/10.1103/PhysRevE.84.036103
5. Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. Nature Reviews Genetics, 5(2), 101-113.
6. Barber, M. J., & Clark, J. (2019). Detecting overlapping community structure in networks. Physical Review E, 90(1), 012805. https://doi.org/10.1103/PhysRevE.90.012805
7. Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2-3), 191-203. https://doi.org/10.1016/0098-3004(84)90020-7
8. 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. https://doi.org/10.1088/1742-5468/2008/10/P10008
9. Erdős, P., & Rényi, A. (2013). On random graphs. Publicationes Mathematicae, 6, 290-297.
10. Fortunato, S. (2016). Community detection in graphs. Physics Reports, 486, 75-174. https://doi.org/10.1016/j.physrep.2009.11.002
11. Fortunato, S. (2016). Extending modularity to overlapping communities. Journal of Statistical Mechanics: Theory and Experiment, 2016(6), 064006.
12. Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826.
13. Gregory, S. (2010). Finding overlapping communities in networks by label propagation. New Journal of Physics, 12(10), 103018. https://doi.org/10.1088/1367-2630/12/10/103018
14. Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. Physical Review E, 83(1), 016107. https://doi.org/10.1103/PhysRevE.83.016107
15. Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. Physical Review E, 80(5), 056117. https://doi.org/10.1103/PhysRevE.80.056117
16. Lancichinetti, A., Fortunato, S., & Radicchi, F. (2009). Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics, 13(3), 033005.
17. Lancichinetti, A., Fortunato, S., & Radicchi, F. (2011). Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics, 13(3), 033005. https://doi.org/10.1088/1367-2630/13/3/033005
18. Lancichinetti, A., Fortunato, S., & Radicchi, F. (2012). Benchmark graphs for testing community detection algorithms. Physical Review E, 78(4), 046110.
19. Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2012). Finding statistically significant communities in networks. PLoS ONE, 7(2), e31439. https://doi.org/10.1371/journal.pone.0031439
20. Meilă, M. (2007). Comparing clusterings by the variation of information. In Learning Theory and Kernel Machines (pp. 173-187). Springer, Berlin, Heidelberg.
21. Meilă, M. (2007). Comparing clusterings—an information based

distance. Journal of Multivariate Analysis, 98(5), 873-895.

22. Newman, M. E. (2016). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577-8582. https://doi.org/10.1073/pnas.0601602103
23. Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. Physical Review E, 69(6), 066133.
24. Newman, M. E. J. (2010). Networks: An Introduction. Oxford University Press.
25. Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 026113.
26. Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043), 814-818. https://doi.org/10.1038/nature03607
27. Peixoto, T. P. (2017). Nonparametric bayesian inference of the microcanonical stochastic block model. Physical Review X, 7(1), 011033. https://doi.org/10.1103/PhysRevX.7.011033
28. 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. https://doi.org/10.1103/PhysRevE.76.036106
29. Ravasz, E., & Barabási, A. L. (2003). Hierarchical organization in complex networks. Physical Review E, 67(2), 026112. https://doi.org/10.1103/PhysRevE.67.026112
30. 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. https://doi.org/10.1073/pnas.0706851105
31. Strehl, A., & Ghosh, J. (2002). Cluster ensembles—a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 3, 583-617.
32. Xie, J., Kelley, S., & Szymanski, B. K. (2013). Overlapping community detection in networks: The state-of-the-art and comparative study. ACM Computing Surveys, 45(4), 43.
33. Xie, J., Szymanski, B. K., & Liu, X. (2013). SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. IEEE Transactions on Knowledge and Data Engineering, 25(2), 253-266. https://doi.org/10.1109/TKDE.2011.198
34. Yang, J., & Leskovec, J. (2012). Community-affiliation graph model for overlapping community detection. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12) (pp. 13-21). https://doi.org/10.1145/2339530.2339537
35. Zhang, H., Yao, S., & Tay, Y. (2019). Overlapping community detection via deep generative model. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 10828-10837. https://doi.org/10.1109/ICCV.2019.01108