International Journal of Applied Social Science
(An International Double Blind Peer Reviewed / Refereed Research Journal of Social Science)
Volume 10 (9 & 10), September & October (2023) : 569-574
Received : 12.08.2023; Revised : 27.08.2023; Accepted : 12.09.2023

RESEARCH PAPER
ISSN : 2394-1405
DOI: 10.36537/IJASS/10.9&10/569-574

An in-Depth Analysis of Overlapping Community Detection Algorithms for Multidimensional Big Data Networks

PARVEEN KAUR
Assistant Professor
Department of Computer Science, Guru Nanak College, Ferozepur Cantt (Punjab) India

ABSTRACT

In this concept, overlapping community detection computations for the multifaceted landscape of multidimensional big data networks are examined in detail. The bulk and increasing complexity of today's network data make traditional community detection techniques inadequate. To determine the similarity of edges from similar layers and cross layers, first apply a generic evaluation of edge behavior inside and between layers. Finally, you may add one more community thickness metric for the multidimensional network and part the dendrogram to get rid of the overlapping communities in these layers. By applying our method to both created and certifiable world datasets, we demonstrate its accuracy in identifying overlapping communities in multidimensional networks. In graph and big data analysis, community discovery is a ubiquitous problem. Locating groups of stationary, interconnected centers with little connections to centers outside the pack is part of it. Identifying communities in large-scale degree networks is an important task in many relevant domains. Finally, certain computations function well on artificial networks, but none of the computations can discern the community structure in real networks. This is a result of the notable communities of the computations being markedly different from the communities defined by the meta-data.

Key Words : Community Detection, Multidimensional Network, MultiComm, Overlapping communities, Big Data

INTRODUCTION

The foundation of modern culture is networks in our rapidly digitizing world. Networks are the underlying structure that underpins and influences our interconnected world, whether it be through social ties, data streams, or the complex linkages that fuel groups and organizations. Big data has brought us an incredible deluge of data that surpasses traditional data structures. Within this vast data landscape, a subset of networks—multidimensional big data networks—stands out as particularly perplexing and difficult to analyze. These networks are unquestionably more complex than their traditional counterparts since they integrate a variety of cooperations, features, and elements.

Community discovery is a primary problem in network analysis, and it is at the heart of comprehending these confusing networks. Community detection looks for clusters of hubs that exhibit reliable availability or meaningful connections in an effort to reveal hidden models and architectures within networks. In spite of this, the conventional methods that have worked successfully in the past fall short when it comes to the intricate details of multidimensional big data networks. Overlapping community detection computations have emerged as a potentially effective solution for handling this test. Due to the inherent complexity of multidimensional networks, these computations allow hubs to coexist with numerous communities.

The goal of this research is to thoroughly examine overlapping community identification computations, with an emphasis on their potential applications and practicality in multidimensional big data networks. We explore the intriguing features and challenges of these networks,

How to cite this Article: Kaur, Parveen (2023). An in-Depth Analysis of Overlapping Community Detection Algorithms for Multidimensional Big Data Networks. *Internat. J. Appl. Soc. Sci.,* **10** (9&10) : 569-574.

providing an overview of state-of-the-art computations. In addition, we provide tailored evaluation metrics and conduct precise studies on real-world datasets to analyze the display of these computations. In order to finally contribute to our understanding of how we might interpret the unpredictable designs that underlie the multidimensional big data networks shaping our interconnected world, we aim to provide significant insights into the capabilities and limitations of current methodologies as well as identify avenues for future research.

Literature Review :

In "Community detection in diagrams" (2010), Fortunato conducts a thorough analysis of community detection methods, focusing on how they might be used in various network environments, such as informal groups, natural networks, and mechanical networks. The work provides an itemized assessment of traditional and modern computations, providing critical perspectives on their benefits, limitations, and emerging challenges in community detection. It has proven to be an invaluable resource for experts seeking a comprehensive understanding of the many techniques and guidelines utilized in this field.

Working with the correlation and evaluation of community detection calculations has been made easier by Lancichinetti and Fortunato's "Community detection calculations: A near examination" (2009). The research provides a basic perspective on the validity of the various calculations by efficiently surveying their exhibition in varied network conditions. Through the presentation of metrics for computation evaluation and the emphasis on the importance of benchmarks, this study has guided subsequent research efforts in benchmarking community discovery techniques. Since then, scientists have adopted these evaluation standards to ensure a more comprehensive and uniform approach to handling computation evaluation, hence increasing the reliability and power of results.

Furthermore, a primary problem in community detection is defining ground-truth communities, which is addressed in Yang and Leskovec's work in "Characterizing and assessing network communities in light of ground-truth" (2012). By using external data as a sort of vantage point, the research suggests a method for objectively evaluating the nature of community detection results. Through an examination of the configuration of established communities with verified real designs, this work provides a fundamental perspective on the challenges of various computations. Since then, the concept of ground-truth-based assessment has become central to community detection research, considering more accurate and substantial assessments of computation execution.

The 2011 publication "Tracking down measurably huge communities in networks" by Lancichinetti and Radicchi offers an important advancement in the evaluation of community structure within networks. The authors suggest a method for identifying measurably crucial communities by comparing the significance of observed community structures with unfavorable models. This work represents the task of irregular possibility and gives a basic factual framework for community identification, enabling the discernment of proof of meaningful designs in complex networks. Since then, this method has influenced the development of more potent and understandable community detection computations.

Mucha *et al.* (2010) explore the evolving concept of communities in unique networks in their paper "Community structure in time-subordinate, multiscale, and multiplex networks". The authors examine the confusing challenges associated with multiplex, multiscale, and time-subordinate networks and suggest a method for identifying community structure in these intricate environments. This work highlights how important it is to be flexible and adaptive when doing community detection calculations because real networks often have transient and multilayer properties. It has prepared for the development of computations that, in due course, can handle the evolving concept of communities.

The 2017 paper "The ground truth about metadata and community detection in networks" by Strip, Larremore, and Clauset provides insights into the use of metadata in community detection. The authors emphasize how important it is to incorporate metadata in order to improve the accuracy of community detection findings. Experts can more easily determine the current reality setting and significance of identified communities by integrating metadata into the inquiry, leading to more informative and notable knowledge. This work emphasizes the need for a comprehensive approach to community discovery that takes associated metadata and network topology into account.

Overlapping Communities in Multidimensional

Networks :

In multidimensional networks, overlapping communities refer to the existence of components or hubs that are located in close proximity to several distinct groups or communities inside a network that has numerous facets. Conventional methods of community discovery in such networks often fall short because they assume that hubs belong only to one community. Nevertheless, in multidimensional networks, constituents may have intricate enrollments and linkages across several community forms, necessitating the development of specific computations that can differentiate and dissect these overlapping communities. This concept is particularly important for today's data analysis because networks can handle various frameworks such as recommendation frameworks, natural networks, or interpersonal organizations. By recognizing overlapping communities, one can gain insights into the complex interactions and examples that occur within these networks. Differentiating and focusing on overlapping communities in multidimensional networks is an active field of research with the goal of revealing hidden patterns and improving our ability to understand intricate structures.

In multidimensional networks, overlapping communities take care of a complicated and subtle aspect of network analysis. In order to fully understand this concept, we should divide it into:

1. **Networks:**Structures known as networks are made up of hubs, also known as vertices, connected by edges, also known as connections. These hubs can address many aspects such as individuals within an interpersonal organization, proteins within a natural network, or web sites within a network of hyperlinks.
2. **Multidimensional Networks:**Conventional networks are typically approached from two perspectives, with hubs and edges depicted on a level plane. However, in multidimensional networks, we also consider additional factors that can deal with different types of relationships, connections, or attributes between hubs. These extra elements may deal with time, types of cooperation, or other relevant network data points.
3. **Communities:**Communities in network analysis refer to groups of hubs that are more closely connected to each other than to the rest of the network. These communities often bear comparison to important bases or bunches within the network, such as utilitarian gatherings in a natural network or buddy bunches in an informal community.
4. **Overlapping Communities:**Hubs are often distributed to a single community in a typical community detection scenario, indicating that they have a location with just one meeting. In any event, hubs are able to coexist with numerous communities at the same time thanks to overlapping communities. This is a more realistic representation of real-world scenarios where objects might have several relationships and fit in with different groups or classes.
5. **Significance:**Understanding the complexity and abundance of links in real networks requires an understanding of overlapping communities. An individual within an informal group, for example, may belong to several communities, such as a gaming club, a professional network, and a community meeting. We wish to consider these covers in order to grasp the essence of their associations.
6. **Challenges:**Due to their increased complexity, distinguishing overlapping communities in multidimensional networks can be challenging. It necessitates the development of specific algorithms that can consider the multidimensional concept of the data and identify hubs across many communities. These computations should also show how hubs can play different roles in different communities, leading to confusing instances of crossover.
7. **Applications:**Applications for comprehending overlapping communities in multidimensional networks can be found in many domains, such as recommendation frameworks, science (protein cooperation networks), informal organization analysis, and on and on. It reveals hidden designs, provides insights into hub functions, and enhances our understanding of how to interpret intricate frameworks.

Communities that overlap in multidimensional networks provide a more sophisticated and reasonable perspective on network analysis by acknowledging that hubs might be a part of several communities and exhibit distinct relationships in different contexts. It is critical to identify and focus on these topics in order to get additional insights into the complex network of interconnected frameworks.

Real-World Networks:

In various disciplines of research, true networks—also referred to as complex networks—are an essential framework for visualizing and understanding the intricate relationships between various substances or hubs. These networks are a valuable resource for understanding the complexity of our globally interconnected environment. The universality of certifiable networks is one of their key characteristics; they span a wide range of domains, including software engineering, science, transportation, sociologies, and finance, to name just a few. Within these networks, edges deal with the relationships, cooperation, or connections between the individual pieces while hubs deal with the connections between them.

The low scale quality of some real networks is a prevalent and important feature. This suggests that while the majority of hubs have a fair number of associations, a small percentage of hubs, sometimes referred to as centers, have an abnormally high number of associations. A power-regulation degree conveyance frequently characterizes this sans scale geography, indicating that the network's trademark scales are not all the same. The presence of highly convincing hubs and increased network robustness against random setbacks are two of this component's wide-ranging effects.

Moreover, certifiable networks share the little world attribute in common. Because of this feature, even in very large networks, the majority of hubs may be reached from another hub in a reasonably small number of steps. The "six levels of partition" theory, which contends that every two persons on Earth are connected to a normal of six middle people through a series of social ties, serves as an example of this concept. True networks are extremely successful at spreading data or effects because of the little world feature.

Another indication of genuine networks is heterogeneity. Hubs in these networks often exhibit distinguishing characteristics, functions, or attributes. In reference networks, for instance, some academic publications receive a great deal of attention and play a crucial role in disseminating information, whilst other papers may have a limited effect. The diversity and specialization of hubs within the network contribute to this variability, resulting in examples of influence dispersion and availability that are not uniform.

Real networks are not static; rather, they are frequently dynamic, meaning that over time, their associations and design change. This distinctive quality is essential for identifying shifts in communications across time. For example, relationships arise through informal groups, and diseases spread through epidemiological networks. Understanding network components and creating expectations or mediations in strong frameworks require a close examination of these ephemeral points of view.

Some prominent examples of certifiable networks include social networks such as Facebook and Twitter, where hubs refer to individuals and edges refer to relationships or partnerships; natural networks, such as protein communication networks, where hubs refer to proteins and edges refer to the real connections between them; transportation networks, such as street grids, where hubs refer to intersections and edges refer to streets; and financial networks, such as exchange networks, where hubs refer to countries or organizations and edges refer to exchange relationships.

The field of real-world network analysis research continues to grow, with applications ranging from comprehending the structure and components of networks to predicting the spread of information or diseases, enhancing transportation networks, generating recommendation engines, and much more. Real-world networks provide an intriguing lens through which to view the intricate relationships, illustrations, behavioural patterns, and flaws present in intricate frameworks across a variety of contexts.

METHODOLOGY

This section provides a brief summary of the various computations used in this review for overlapping community detection. In order to do some of these computations, we must define clear limits. The bounds, which we selected for each computation, are found in the attached. A few computations are included in none of these courses.

A few designed benchmark networks and real networks with ground-truth communities were used for the trials. We use artificially generated benchmark networks with small (5000 hubs) and large (50000 hubs). To generate the designed benchmark networks, we employ three distinct models: the Erdös-Renyi [ER59] model as control, the LFR [LFR08] model, and the CKB [CKB+14] model.

We also make use of another type of networks, known as CKB networks [CKB+14]. The CKB benchmark generator provides information on the number of communities a hub is located in in addition to the power regulation circulation of community measures. With the exception of a larger number of least communities, the boundaries for the CKB networks shown in Table 1, Fig. 1, Table 2, and Fig. 2 are chosen in accordance with the concepts of [SHW17]. These boundaries are identical to those suggested in the initial study. Take note that a power regulation conveyance with kinds 2.5 is observed by both community sizes and communities per hub. These CKB networks were constructed using the execution provided

Table 1: Participation in the mix parameter

Description	Mix. Parameter	Membership
M	Number of nodes	121
VNV	Average degree	89.34
K	Max degree	57.2
M	Min. coomm, size	22
WAH	Degree	51.4
OL	Comm., nodes	25.2
NC	Mixing	55.2
T1	Overting	78.2
T2	Comms., per noder	57.3

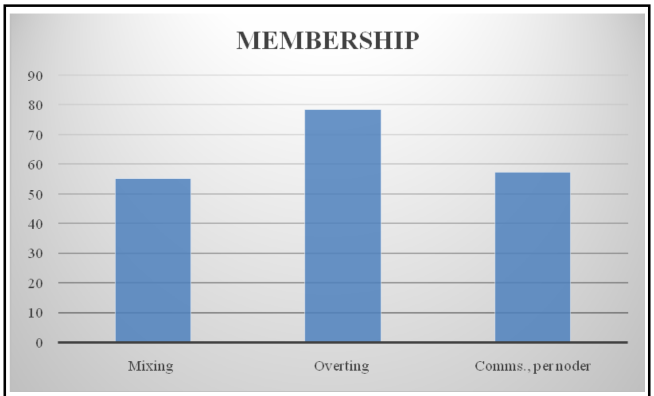

Fig. 1 : Schematic for membership

Table 2 : An explanation of the mix parameter for the large and small CKBs.

Description	Mix. Parameter	Small CKB	Large CKB
N	Number of nodes	50000	5000
M	Average degree	2	1201
XMIN	Max degree	201.2	465
XMAX	Min. choom, size	48.4	25.03
XMIN	Degree	2004	1161
XMAX	Comm., nodes	1641	14.23

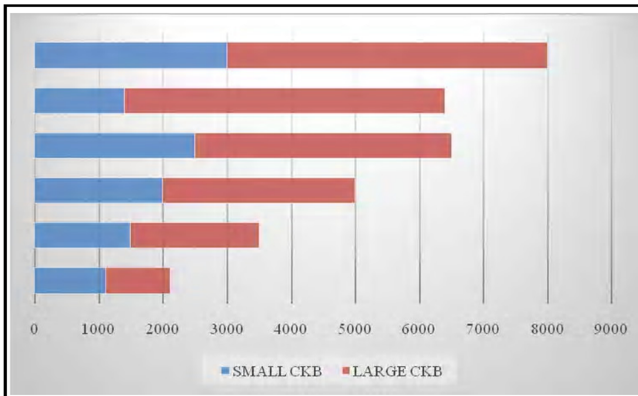

Fig. 2 : Diagram illustrating the mix parameter definition for the small and large CKB

by [SHW17]. Furthermore, we employ irregular networks, namely Erdös-Renyi networks [ER59]. The Erdös-Renyi network generator just requires two parameters: the number of hubs and the edge likelihood. Each vertex set is then probabilistically correlated with every other set of vertices. This creates a network that shouldn't have any kind of community organization.

RESULTS AND DISCUSSION

The results of the overlapping community detection estimations on different designed and certifiable networks are described and shown in this section. Ten instances of each boundary configuration were generated for designed networks, and each computation was performed many times for real networks. We allow for a maximum run season of four hours in each computation. The initial stages were conducted on a server equipped with a quad core Intel processor (Intel Centre i7-2600 K CPU chip operating at 3.40 GHz), 32 GB of Smash, and hyperthreading enabled.

Conclusion:

Overall, our comprehensive analysis of multidimensional big data networks' overlapping community detection computations has shed light on the complexity and challenges inherent in this rapidly developing subject. These analyses have not only helped us better understand the various approaches available for breaking down multidimensional networks, but they have also provided fundamental evaluation metrics, benchmarking procedures, and considerations for multiplexity, metadata, and time dependability. As a result, our study has provided a solid foundation for further research and the development of more accurate and careful computations, ensuring that we will be able to continue exposing the complex structures and components found inside multidimensional big data networks in a world that is unquestionably interconnected.

REFERENCES

Ball, Brian, Karrer, Brian and Newman, M.E.J. (2011). An efficient and principled method for detecting communities in networks. April 2011.

Dhanya, Sudhakaran and Shini, Renjith (2015). Phase based resource aware scheduler with job profiling for Map Reduce, *IJLTET*, **6** (2) : 92-96.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**(3-5) : 75-174.

Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, 80(5), 056117.

Lancichinetti, A. and Radicchi, F. (2011). Finding statistically significant communities in networks. *PloS ONE*, **6**(4), e18961.

Leskovec, J., Lang, K.J. and Mahoney, M.W. (2010). Empirical comparison of algorithms for network community detection. CoRR, abs/1004.3539, 2010.

Moradi, F., Olovsson, T. and Tsigas, P. (2012). An of community detection algorithms on large-scale email traffic, in: SEA. Berlin/Heidelberg: Springer; 283–294.

Mucha, P.J., Richardson, T., Macon, K., Porter, M.A. and Onnela, J.P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, **328**(5980) : 876-878.

Newman, M. (2003). The Structure and Function of Complex Networks," *SIAM Rev*., **45** (2) : 167-256.

Nguyen, Nam P., Dinh, Thang N., Xuan, Ying and Thai, My T. (2011). Adaptive Algorithms for Detecting Community Structure in Dynamic Social Networks", IEEE infocom.

Pascal Pons and Matthieu Latapy (2005). "Computing communities in large networks using random walks. *ISCIS, Springer,* **3733** : 284- 293.

Peel, L., Larremore, D.B. and Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances*, **3**(5) : e1602548.

Raghavan, Usha Nandini, R eka Albert and Kumara, Soundar (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, **76**(3) : 036106.

Renjith, Shini and Anjali, C. (2013). Fitness function in genetic algorithm-based information retrieval: A Survey", ICMIC13, December 2013, pp. 80-86.

Riedy, Jason, Karl, A. David Bader, Pande, Jiang Pushkar and Sharma, Richa (2011). Detecting Communities from Given Seeds in Social Networks, February 22, 2011.

Tang, Lei, Wang, Xufei and Liu, Huan (2010). Community Detection in Multi-Dimensional Networks", Technical Report, TR-10-006, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287.

Wei Chen, Zhenming Liu, Xiaorui Sun, Yajun Wang (2010). A game-theoretic framework to identify overlapping communities in social networks. *Data Min Knowl Disc,* **21**:224–240.

Xiaoming Liu, Yadong Zhou, Chengchen Hu, Xiaohong Guan and Junyuan Leng (2014). Detecting Community Structure for Undirected Big Graphs Based on Random Walks", www' 14 Proceedings of 23rd international conference on WWW, 2014, pp.1151-1156.

Yang, J. and Leskovec, J. (2012). Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, **42**(1) : 181-213.

****