ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 12 December 2023

Network Analysis in Big Data Research

[1] S.S.Raja Kumari, [2] Dr. CH.Nagaraju *

[1] Associate professor of CSE Department, St.Johns College of engineering and technology,Yerrakota, Yemminganur, Kurnool district, AP-518360,

[2] Professor and HOD, Dept of ECE, Annanamacharya institute of technology and sciences (AITS), Rajampet- 516126

Email: [1] ssrajakumari2009@gmail.com, [2] chrajuaits@gmail.com

***Abstract*— The use of network analysis in big data research looks into the significance of studying trends in communication among nodes and between groups. Big data offers the potential to change a variety of companies, but accomplishing this effectively requires knowledge in fields as diverse as computer science and mathematics. The literature overview covers practical issues, challenges, and conceptual underpinnings. The theoretical underpinnings of social network theory are also investigated in this study. It provides results in the recognition of community algorithms and the framework of big data networks. The research shows the growing importance of employing network analysis in the constantly shifting big data context.**

***Index Terms*— Network Analysis, Big Data Research, Connectivity, Community Trends, Interdisciplinary, Computer Science, Mathematics, Optimization, Cloud Computing.**

I. INTRODUCTION

"Network analysis (NA)" is a group of techniques that plot connections between individuals and analyse the trends within communities that emerge as a consequence of these connections developing repeatedly. The basic idea is that studying the relationships between objects is essential to comprehend social events. Network analysis could be applied to a social network to determine which people are most effective in terms of propagating an idea or behaviour. Bottlenecks and other possible sources of congestion in a transportation network can be located using network analysis [1]. Different industries have changed as a result of big data, which is becoming more common due to the proliferation of Internet-connected and portable devices. Knowledge from several fields, especially CS and maths is required for the methodical growth of big data and big data analytics. Many research projects on big data, particularly social media information have been conducted in a wide variety of industries as a result of the increasing interest in analytics and big data.

II. LITERATURE REVIEW

Applications of Network Analysis in Big Data Research

Optimization and the administration of networks are essential applications for big data in this setting. Network administrators may profit from big data analytics by monitoring traffic, seeking outliers, and adjusting settings for optimal performance. A great deal of data can be analysed in real time, enabling network administrators to swiftly identify problems and take preventative measures to prevent interruptions in service [2]. In addition, by analysing both past and present information, big data analytics can direct network managers toward better investments, abilities, and quality decisions. Enterprises may utilize cloud computing by storing data, such as subsurface monitoring data and underground production rates, that would otherwise require extraordinarily enormous storage equipment.

Figure 1. Necessity of big data analytics

As new technologies in computation and network connectivity form the basis, the current cloud computing instructor technology develops. Its main use is the processing and storage of huge amounts of data, giving it an edge over the competition when it involves cloud computing installations in companies of any size [3]. Detection as well as early warning are essential stages in the process of manufacturing. Enterprises need to gather monitoring information, such as the ambient temperature, moisture, flow of air, and other kinds of facts in the company's factories and machinery. It includes real-time data from multiple machines, in order to accumulate a great deal of data to identify the manufacturing process of the internal environment.

Challenges and Limitations in Network Analysis of Big Data

Collecting and processing the enormous quantity of data associated with NTA is an important challenge. NTA could

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 12 December 2023

require terabytes as well as petabytes of storage and computational resources, according to the size and design of the network. In addition, data could have different protocols, encryption stages, and types of files as it flows from different gadgets, including switches, routers, firewalls, and terminals. This limits the capacity to do accurate and timely correlation and evaluation. Protecting the confidentiality and integrity of the data is a further challenge for NTA. Depending on the type and nature of the data, NTA could involve critical or private data, such as IP addresses, identities of users, or emails [4]. Especially if data traverses international or organizational boundaries, this might create legal and moral problems.

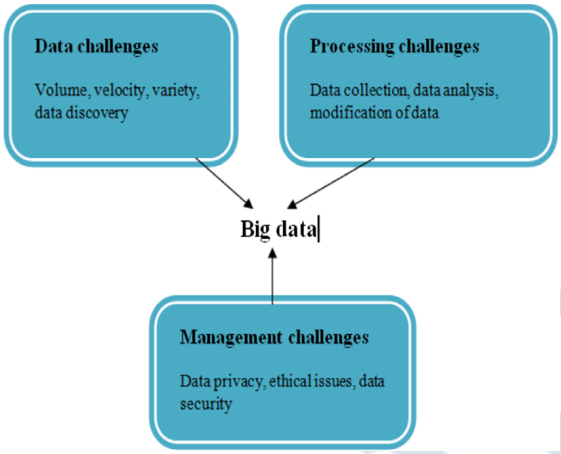

Figure 2. Big data challenges

Therefore, NTA professionals have to conform to applicable laws and regulations, such as the "General Data Protection Regulation (GDPR)" and the "Health Insurance Portability and Accountability Act (HIPAA)", and carry out suitable safeguards such as anonymization, encryption, or permission to safeguard the information from unlawful use or misuse. Furthermore, NTA faces difficulty in gathering and evaluating data of superior quality [5]. Due to problems such as congestion in the network, dropped packets, or set-up errors, the data could become ineffective, incorrect, or deficient in vital information.

Theoretical Foundations and Frameworks in Network Analysis

Social Network Theory and Analysis

Using the conceptual frameworks of sociology and the mathematical foundations of graph theory, the analysis of social networks examines the connection between architecture and health. The majority of investigations of social networks either depend on concentrating on their entirety that is socio-centric or on specific participants which is ego-centric [6]. Whole-network investigations analyse relationships between individuals or groups that, for analytical simplicity, are presumed to exist within an organized framework, even if those boundaries are dynamic or ambiguous.

The computations of descriptive metrics, which reveal important features about the standing of network actors, the features of network groups, and the features of whole networks, are the core of social network analysis, whereas visualizations of information are useful. As an indicator of network unity, the position of participants in the network or the extent to which they have connections is usually used.

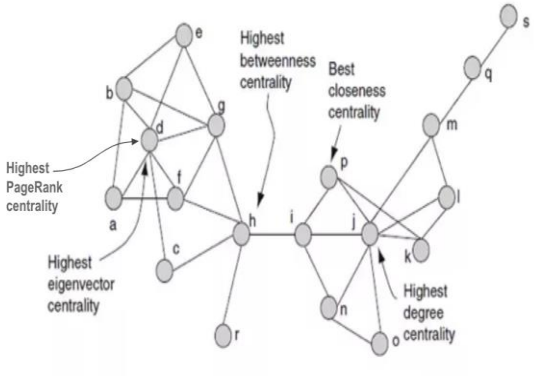

Figure 3. social network theory and analysis

III. METHOD

Constructivism research philosophy has been chosen for the present study. A philosophical paradigm referred to as constructivism claims that there are multiple realities as everyone constructs a unique interpretation of reality using their thoughts. The interpretation and evaluation of big data can be extremely personal, especially when interacting with complex networks. It aligns perfectly with the theory that the assessment of patterns and connections in big data networks can be highly subjective and contextual.

The researchers decided that a longitudinal strategy would be best for this type of study. Development and transformation are common in big data networks. Researchers may track networks over the years and observe their ability to grow and change via a longitudinal study design [7]. Understanding the constantly shifting nature of big data networks is of the utmost significance. Secondary qualitative data analysis method has been selected for conducting the entire study. The data has been collated from articles, journals from google scholars and different websites also have been used to collect data. Moreover, the collected data has been analysed by thematic analysis method. This method has helped to make well-structured and descriptive method of data analysis.

IV. RESULT

Theme 1: Community Detection and Clustering in Big Data Networks

The strategy includes developing an innovative type of community discovery algorithm termed as "Hyperbolic Girvan-Newman (HGN)" [8]. In addition to its application for data clustering, HGN can also be used with data graphs

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 10 Issue 12 December 2023

for pure population discovery, such as social networks. In HGN, a restricted number of landmarks are used to set the graph within hyperbolic space using the Rigel embedding. The landmarks serve as anchor nodes, from which distances to the other embedding vertices are determined. As compared to the distances of each of the nodes to the destinations, the ones calculated among the landmarks themselves tend to be more precise [9]. Algorithm 1 follows to figure out the HEBC score for every edge. A predetermined number of connections with the highest EBC values, called the group, is selected according to the notion that such edges are most likely to operate as "bridge-like" relationships, bringing together two potential communities, and that such connections are a small proportion of the total.

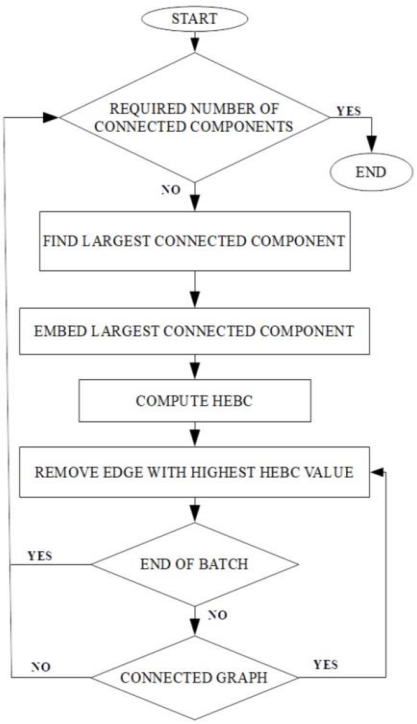

Figure 4. Hyperbolic Girvan–Newman algorithm.

Theme 2: Architecture of big data networking is important network analysis

Hadoop Apache's central part, MapReduce is a programming model to handle large datasets that may be developed using a variety of languages, including Java and C++ [10]. Through breaking up petabytes of records into smaller pieces, MapReduce has the ability to handle them in parallel on Hadoop commodity machinery. Furthermore, it compiles data taken from many servers and provides an aggregate report back to the customer's program. This application can break down larger applications into smaller portions that can subsequently be distributed to individual nodes in a network topology. In order to filter, sort, and redistribute jobs to various nodes and then collect the outcomes of those tasks into one number for quick analysis, it uses a map method.

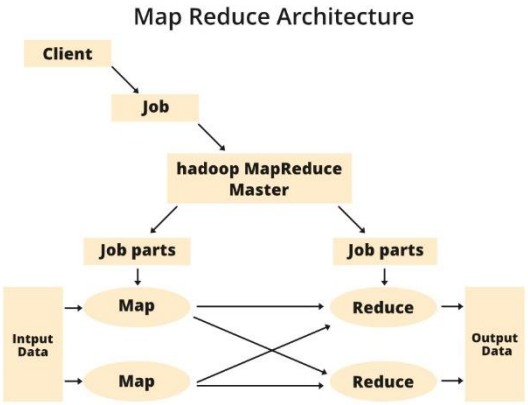

Figure 5. Architecture of MapReduce

On the other hand, Hadoop is an administration software platform, which is crucial to the discipline of big data analytics. Hadoop was developed for application in a distributed system setting, where it is able to rapidly manage, organize, share, and analyse big, unorganized data sets among many nodes [11]. Data is stored in Hadoop's HDFS, a file system that is distributed and saves data in blocks over a cluster of machines. Hadoop and MapReduce are among the most popular tools for building big data infrastructure at present [12]. The client serves as the main point of contact and search engine. Information is preserved in nodes with localized databases, which then spread across a network horizontally. The distributed task's components are handled by the task tracker once each node gets its share. In order to determine which data node holds the desired data, the name node maintains a position index of all other nodes in the system.

V. CONCLUSION

The significance of network analysis to the discipline of big data research is highlighted. Complex networks and implications across sectors are highlighted, along with applications, issues, and theoretical underpinnings. Advances in data management and strategic choices are provided by studies on detecting community algorithms and the design of big data networking. This study shows the increasing significance of using network analysis in the constantly shifting big data system.

REFERENCES

[1] P. Chunaev, "Community detection in node-attributed social networks: A survey," Computer Science Review, vol. 37, p. 100286, 2020.

[2] "The evolution of social network theory: Perceived impact on developing networking relationships," American Journal of Management, vol. 19, no. 3, 2019.

[3] "The evolution of social network theory: Perceived impact on developing networking relationships," American Journal of Management, vol. 19, no. 3, 2019.

[4] I.-L. Hu, C.-C. Chang, and Y.-H. Lin, "Using big data and social network analysis for Cultural Tourism

Planning in Hakka villages," Tourism and Hospitality Research, vol. 21, no. 1, pp. 99–114, 2020.

[5] M. Ianni, E. Masciari, and G. Sperlí, "A survey of Big Data Dimensions VS social networks analysis," Journal of Intelligent Information Systems, vol. 57, no. 1, pp. 73–100, 2020.

[6] D. Jin et al., "A survey of Community Detection Approaches: From Statistical Modeling to Deep Learning," IEEE Transactions on Knowledge and Data Engineering, pp. 1–1, 2021.

[7] X. Li and R. Law, "Network analysis of Big Data Research in Tourism," Tourism Management Perspectives, vol. 33, p. 100608, 2020.

[8] N. Moustafa, "A systemic IOT–fog–cloud architecture for big-data analytics and cyber security systems," Secure Edge Computing, pp. 41–50, 2021.

[9] A. G. Polyakova, M. P. Loginov, A. I. Serebrennikova, and E. I. Thalassinos, "Design of a socio-economic processes monitoring system based on network analysis and Big Data," International Journal of Economics and Business Administration, vol. VII, no. Issue 1, pp. 130–139, 2019.

[10] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of Bridge Structural Health Monitoring aided by Big Data and artificial intelligence: From condition assessment to damage detection," Journal of Structural Engineering, vol. 146, no. 5, 2020.

[11] Y. Tao, A. Petrović, and M. van Ham, "Commuting behaviours and subjective wellbeing: A critical review of longitudinal research," Transport Reviews, vol. 43, no. 4, pp. 599–621, 2022.

[12] F. Li, K. Zhang, S. Chen, H. Yang, and B. Wang, "Research on key technologies of active defense for distribution internet of things service security," 2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA), 2020.