Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.032127
Article

Tech Science Press

Blockchain Based Consensus Algorithm and Trustworthy Evaluation of Authenticated Subgraph Queries

G. Sharmila[1,*] and M. K. Kavitha Devi[2]

[1]Department of Computer Science & Engineering, Syed Ammal Engineering College, Ramanathapuram, 623502, India
[2]Department of Computer Science & Engineering, Thiagarajar College of Engineering, Madurai, 625015, India
*Corresponding Author: G. Sharmila. Email: gsharmilag89@gmail.com
Received: 07 May 2022; Accepted: 10 June 2022

Abstract: Over the past era, subgraph mining from a large collection of graph database is a crucial problem. In addition, scalability is another big problem due to insufficient storage. There are several security challenges associated with subgraph mining in today's on-demand system. To address this downside, our proposed work introduces a Blockchain-based Consensus algorithm for Authenticated query search in the Large-Scale Dynamic Graphs (BCCA-LSDG). The two-fold process is handled in the proposed BCCA-LSDG: graph indexing and authenticated query search (query processing). A blockchain-based reputation system is meant to maintain the trust blockchain and cloud server of the proposed architecture. To resolve the issues and provide safe big data transmission, the proposed technique also combines blockchain with a consensus algorithm architecture. Security of the big data is ensured by dividing the BC network into distinct networks, each with a restricted number of allowed entities, data kept in the cloud gate server, and data analysis in the blockchain. The consensus algorithm is crucial for maintaining the speed, performance and security of the blockchain. Then Dual Similarity based MapReduce helps in mapping and reducing the relevant subgraphs with the use of optimal feature sets. Finally, the graph index refinement process is undertaken to improve the query results. Concerning query error, fuzzy logic is used to refine the index of the graph dynamically. The proposed technique outperforms advanced methodologies in both blockchain and non-blockchain systems, and the combination of blockchain and subgraph provides a secure communication platform, according to the findings.

Keywords: Big data; blockchain; consensus algorithm; trust management; graph index

1 Introduction

In the field of data mining, graph mining has been a popular topic. It is now possible to represent a large amount of data with a graph. There are several application fields based on graph mining and some of the fields are as follows: (1). Chemical informatics, (2). Bioinformatics, (3). Computer vision, (4). Video indexing, (5). Text retrieval, (6). Web analysis and (7). Social networks [1–4]. In this field, graph index

construction has been paid great attention, which aims to speed up the query processing [5]. The difficulties produced by urbanization, such as traffic jams, environmental degradation, a lack of resources, and a reduction in people's quality of life, have grown increasingly evident as cities' populations continue to grow and new urban agglomerations arise. The notions were introduced to achieve cities' long-term sustainability [6]. An important aspect of graph theory is finding patterns in graphs' subgraphs, which can also be known as network motifs or graphlets.

Making more efficient decisions in a blockchain requires the effective processing and utilization of big data [7]. In this field, graph index construction has been paid great attention to, which aims to speed up query processing. These approaches are very computational and cost-effective. Graph indexes must act as a filter to prune the false candidate set from the resulting answer set at a reasonable cost [8]. Fig. 1 shows the background of subgraph mining [9]. Currently, query service providers are untrusted and compromised by attackers. Authenticated Query Search is invoked as a new idea in graph mining [10]. In this idea, two properties are important to meet i.e., Quality of service (QoS) and Security. Blockchain is a new technology that provides security for the decentralized environment. This will completely prohibit the generation and tampering with the result of the answer set from the query service providers.

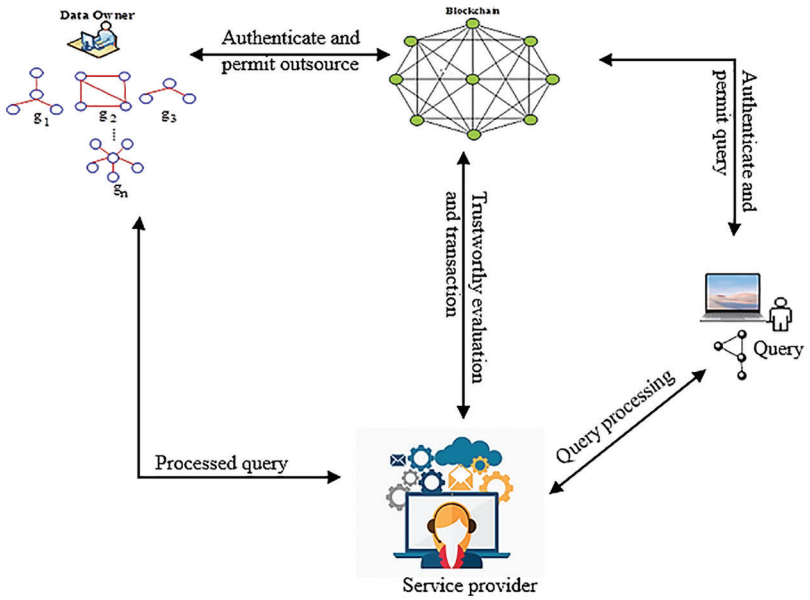

Figure 1: Subgraph mining background

In the current digitalized world, users are working in the distributed environs i.e., the size of graph databases are very huge since graph mining emerged in various application sources [11]. In the centralized approach, the number of scans to mine the candidate answer set requires more time and it is cost-effective because it uses traditional filtering and verification framework, insufficient feature set extraction, pivot element selection, subtree patterns finding, and so on [12]. In addition, sub-graph isomorphic testing does not meet all the constraints from the graph index. The blockchain layer is responsible for data computing and comprehensive data processing.

The most critical problems in subgraph mining have been subjected to, less data collection, excessive query processing time due to poor index construction, feature selection and isomorphic testing, insufficient large scale datasets and the Absence of semantic (context) relationship between query graph and graph database, it can also suffer from security issues. To overcome these problems in graph

indexing and authenticated query processing, the Blockchain-based Consensus algorithm for authenticated query search in the Large-Scale Dynamic graphs (BCCA-LSDG) technique is proposed.

A Blockchain-based mechanism is used for assessing the trustworthiness of the subgraph query. The trustworthiness evaluation cannot be completed by itself due to the restricted computational and storage capacity. The trustworthiness of blockchain is assessed using reputation, which takes into account powerful processing resources and storage capacity. Blockchain has shown to be a novel way to security issues. It is a distributed network that uses a consensus method to ensure multiple nodes agree on the status of particular data. Blockchain technology will expand the digital network of internet-based technologies, revolutionize it, and automate it. Consensus algorithms ensure that the blockchain network is stable and that unidentified peers have faith in the distributed computing system. The consensus mechanism ensures that a new block added to the blockchain is the valid iteration that has been agreed upon by all blockchain peers. In this paper, two-fold processes are implemented such as graph indexing and authenticated query processing with the improvement of user accuracy as well as privacy protection for owners. The overall scheme is implemented using a Hadoop environment which is investigated by conducting three analyses such as storage, security, and scalability. The incorporation of Hadoop and blockchain ensures security as well as performs faster. The Map Reduce and machine learning algorithms used are effectively presented in this paper.

The significant contributions of the BTLA-LSDG are summarized as follows,

- In this paper, two-fold processes is implemented such as graph indexing and authenticated query processing with the improvement of users accuracy as well as privacy protection for owners.
- Firstly, data owner is authenticated to blockchain using a consensus algorithm, which generates session token and secret key for owners.
- For graph indexing and data storage, Trustworthy Evaluation Module is used.
- We use MapReduce process with the Structural Similarity and Semantic Similarity for the graph and query so it runs in a parallel mode. It requires very less amount of time.
- The proposed work with Hadoop MapReduce is also supported for faster retrieval which is enabled to support enormous numbers of users with accurate subgraph retrieval.
- The overall scheme is implemented using Hadoop environment which is investigated by conducting three analysis such as storage, security, and scalability. The incorporation of Hadoop and blockchain ensures security as well as it performs faster.

The remaining section of this research is demonstrated as follows. The literature review is explained in Section 2. The suggested techniques were shown in Section 3, along with an explanation and the related algorithm. The performance outcome and their analysis are provided in Section 4. Conclusions and further work were completed in Section 5.

2 Literature Review

In 2019, Zhaofeng et al. [13] proposed a Block BDM (Blockchain-based Big Data Management) to address the security and trust issues associated with Internet of Things (IoT) data management. As well as safe use controls for digital rights management, they offered a token-based solution for high-value data consumption to prevent theft and distribution without authorization. A large amount of testing revealed that the suggested Block BDM method for decentralised trust management of IoT big data is viable, secure, and scalable.

In 2020 Chae et al. [14] proposed a Dynamic feature selection for Bag Classification (DBC) algorithm for classifying bags incrementally. DBC uses incremental gScore and incremental gSpan as a core part of its algorithm. According to our experiments, our method produces a significantly more informative feature set than the one that was originally designed for static big data with less accuracy loss.

In 2021, Pampapathi et al. [15] proposed an Improved Adaptive Neuro-Fuzzy Inference System (IANFIS) and Modified Elliptical Curve Cryptography (MECC) in IoT which provide effective information sharing and safe data communication. A cloud server receives the sensed data from numerous IoT devices through IANFIS. Through the MECC, the user receives the data securely. IANFIS' proposed classifier achieves a 311 millisecond response time for 5000 data points while achieving 96 percent security.

In 2021 Qureshi et al. [16] proposed the Trust Evaluation Model for Smart Grids (TEMSG) to secure the data collection in smart cities and smart grids. In smart grids, machine learning algorithms can be used to estimate faulty information and, consequently, to obtain trust values. Experiments are carried out to assess and evaluate the suggested framework of accuracy, reliability, and detection rate.

In 2021, Yadav et al. [17] proposed a platform based on the blockchain for property transaction digitization. A consensus mechanism is suggested to make it secure, and it decreases the transmissions for the multicasting nodes by around 50%. In contrast to a standard PoW strategy in which all nodes engage in consensus, the overhead for message exchange communication is cut by 54.87 percent, and the time taken for consensus is reduced by approximately 53.7 percent and 10%.

In 2021, Asefi et al. [18] proposed a combined blockchain-based communication platform and distributed state estimation for safer and more stable data transmission Furthermore, the asynchronous mode of data transfer has been examined as the secondary aim of this study, which is more likely to occur in the actual world. The numerical analysis reveals that the suggested strategy improves the distributed state estimation process' security and reliability.

In 2022 Cong et al. [19] proposed a privacy-preserving subgraph isomorphic query over graphs. In addition, these approaches depend on identifying specified subgraphs as index characteristics, which imposes significant computational burdens on the graph owner. Additionally, they must search the entire index to find data graphs that match the criteria, causing the query to be slow. According to the performance evaluation and the security proof, this technique is both efficient and secure.

In 2022 Sun et al. [20] proposed a subgraph matching algorithm based on the subgraph index for FGqT-Match. Subgraph matching involves two key designs for indexing subgraphs. Matching-driven flow graphs (FGqT) are designed to eliminate superfluous calculations in advance through subgraph indexes. Research on actual and synthetic graphs shows that this solution outperforms the existing method.

In 2022 Wang et al. [21] proposed the Reinforcement Learning (RL) and Graph Neural Networks (GNNs) method to create the high-quality matching order for the subgraph matching method. Reinforcement Learning Based Query Vertex Ordering Model (RL-QVO) learns the strategy for generating the matching order by taking into account both the graph structure and the long-term advantages. In comparison to existing methods, our proposed matching order generating strategy reduces query processing time up to two orders of magnitude when applied to six real-world data graphs.

The above-reviewed methodologies possess some drawbacks in sub graph mining such as several digital security threats, query service providers are untrusted and compromised by the attackers, scalability due to insufficient storage, security threats. To overrides these problem the Blockchain-based Consensus algorithm for Authenticated query search in the Large-Scale Dynamic graphs (BCCA-LSDG) technique is proposed. In addition, we considered the "storage optimization", which was not focused at all in previous works. By using consensus algorithm the data is authenticated in a secure manner. The trustworthy module is constructed using address based blockchain reputation system to ensure the original data from data poisoning attack. Then Dual Similarity based MapReduce helps in mapping and reducing the relevant subgraphs with the use of optimal feature sets. The proposed BCCA-LSDG scheme has unique merits compared with existing methods in terms of data management and query processing since these are advanced factors in this field.

3 Proposed Methodologies

The functionality of each component employed in the creation of the proposed framework, including the trust evaluation model and secure massive data transmission, is described in this section. The Consensus technique is used to achieve dependability in a network with several unreliable nodes in the secure big data transmission module. The entire process handled in this proposed system is depicted in Fig. 2. Firstly, the data owner is authenticated to the blockchain using a consensus algorithm, which generates a session token and secret key for owners.

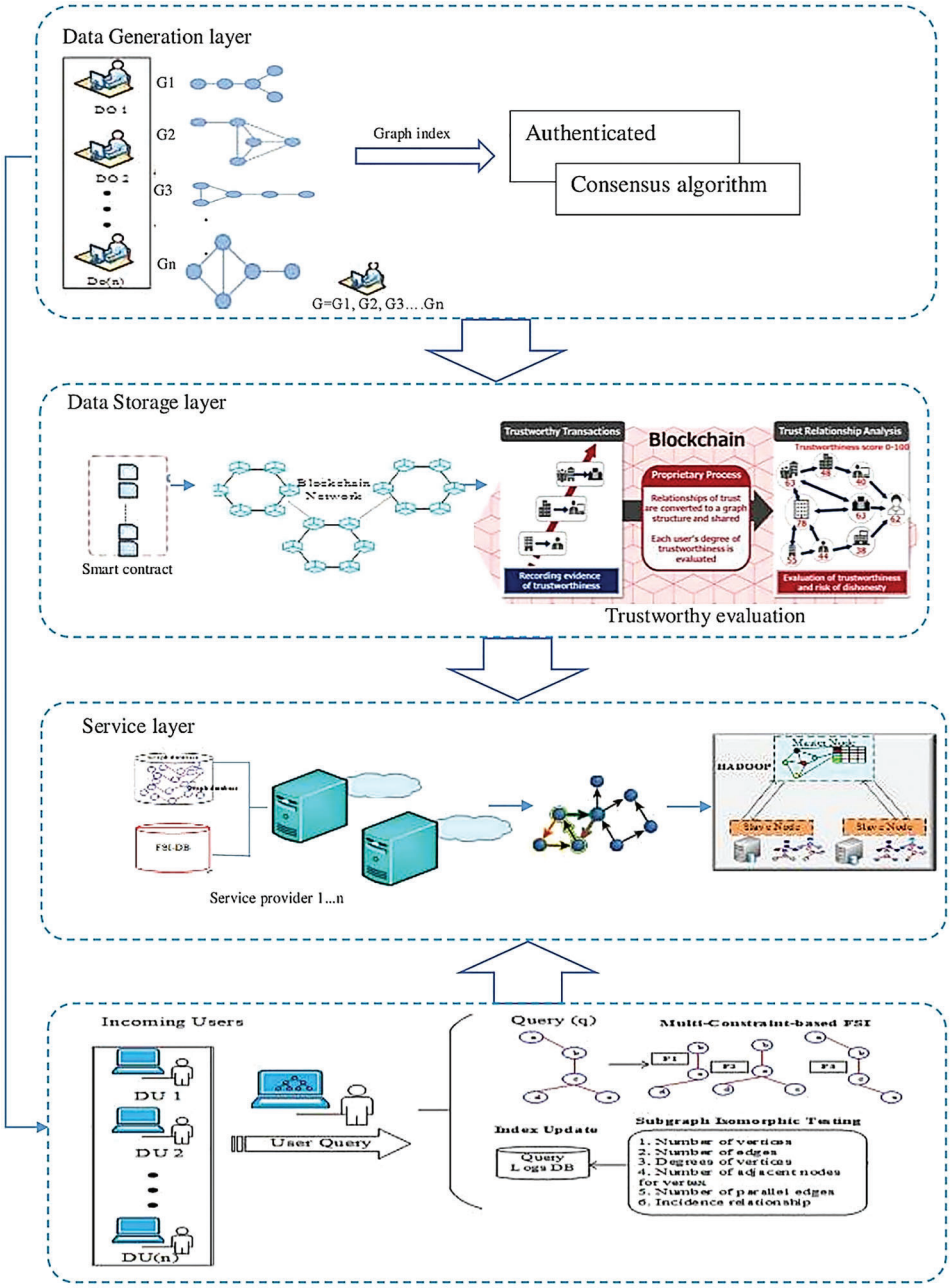

Figure 2: Proposed architecture

3.1 Data Generation

Authentication: Consensus Algorithm

The blockchain is initially authenticated by consensus algorithms. As shown in Fig. 3, a blockchain-based secure data transmission flow chart shows how secured data is transmitted. Different blockchains are categorized according to the data they handle, the services they provide, and the behaviors they allow their users to engage in. Data can be separated between public and private blockchains, as well as permissionless and permissioned blockchains. Users are able to browse public blockchains and validate transactions without requiring permission from third parties. Some decentralised blockchains, such as Bitcoin, offer economic incentives to miners and validators. In private blockchains, network connectivity is restricted by the controller. All private blockchains must automatically monitor which operators are allowed to move data and execute smart contracts. The three most common consensus techniques for shared blockchains are Proof of Stake (PoS), Proof of Work (PoW), and Delegated Proof of Stake (DPoS).

Determine X_i the possibility for a malignant user to be added to the equivalent list in the event of blocks falling behind. The data on the Y-axis may change the unit destination each time, with the prospective for X to the right or Y to the left (X + Y = 1).

The node is initially at X = Y, if the particle reaches Q = 0, it will stop spinning. The probability of arriving at Q = 0 is equal to X_i. So:

$$X_\varphi = 1,\ lim_{H-\to\infty} X_i = 0 \ldots \tag{1}$$

From Eq. (1) H denotes the total hashing power of the network.

$$X_i = \mu X_{i+1} + YX_{i-1},\quad where\ i = 1,\ 2,\ 3 \ldots \infty \ldots \tag{2}$$

where X_i denotes the probability of the malicious nodes. X denotes the hashing rate of the honest nodes. Y denotes the total hashing rate of malicious nodes.

If Y < X, use this equation

$$Z_i = X_{i+1} - X_i = M = \frac{Y}{X} \ldots \tag{3}$$

From Eq. (2)

$$X_i - X_\varphi = \sum_{n=0}^{i-1} (X_{n+1} - X_n) = \frac{1 - M^i}{1 - M} Z_\varphi \ldots \tag{4}$$

Eq. (1) is therefore written as

$$X_i = M^i = \left(\frac{Y}{X}\right)^i,\quad where\ Y < X \ldots \tag{5}$$

It is easy to prove that

$$X_i = 1,\quad where\ Y >= X \ldots \tag{6}$$

The last equation is,

$$X_i = \begin{cases} 1, & Y >= X \\ \left(\frac{Y}{X}\right)^i, & Y < X \end{cases} \cdots \tag{7}$$

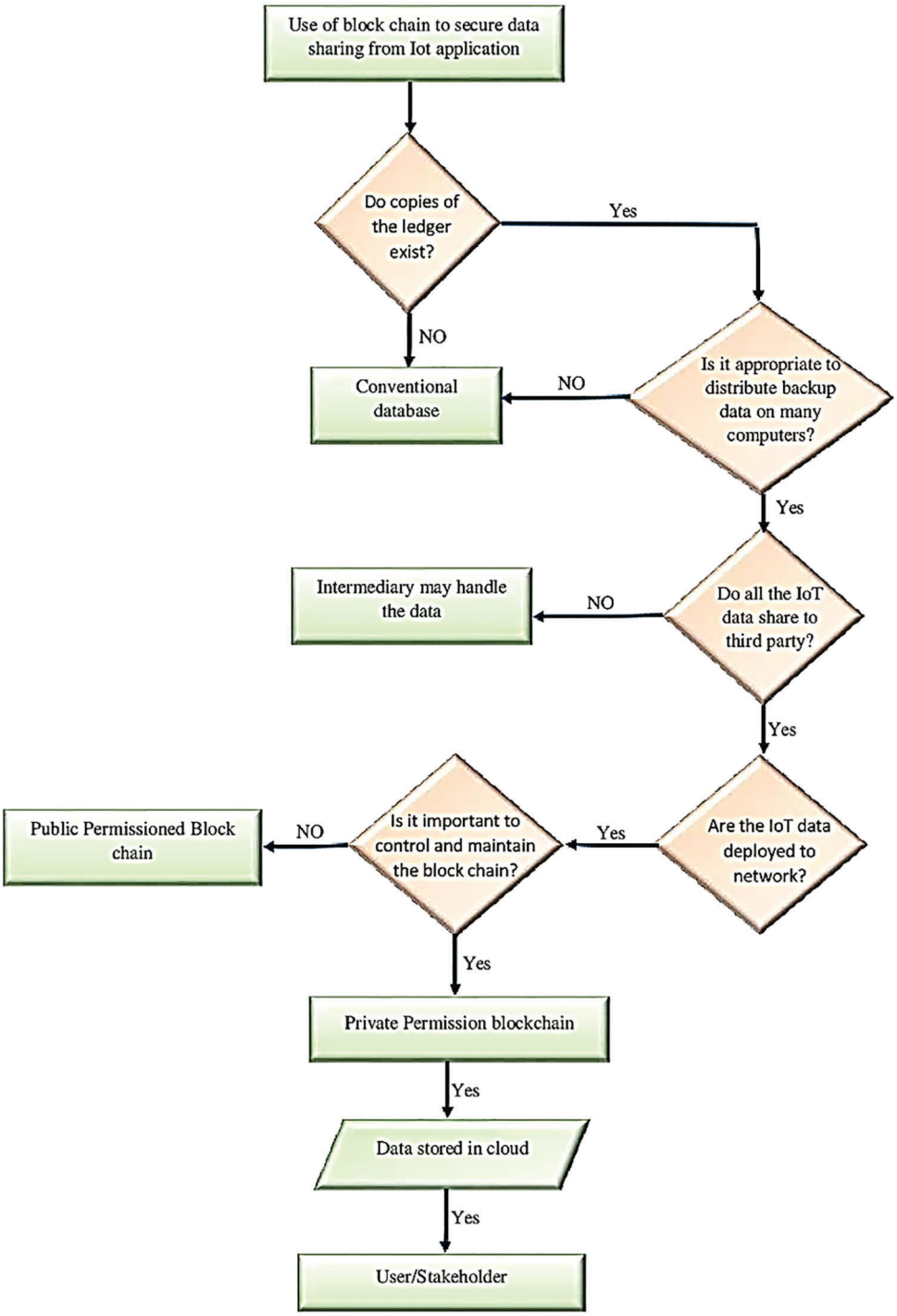

Figure 3: Flowchart for consensus algorithm for authentication and sharing

Double-spending attacks succeed when the malicious program's capacity exceeds the network's capacity. Determine if i blocks will reach double expenditures before YX. When the malicious nodes require double-spending attacks, wait until the data transport is checked for i blocks.

In the event that a node is unable to find the correct hash within t seconds, it will restart hashing. Honest nodes have X/T_0 chance of success, while malicious nodes have Y/T_0 chance. T_o represents the average time for creating a new block.

To perform a double-spending attack, malicious nodes must wait n blocks for the transactions to be validated. Poisson distributions of predicted values characterize the effective timings of malicious nodes.

$$\gamma = iY/A \ \ldots \tag{8}$$

So the likelihood of the malicious nodes' successful double-spending is written as;

$$X = 1 - \sum_{n=0}^{i} \frac{\gamma^n \epsilon^{-\gamma}}{n!} \left[1 - \left(\frac{Y}{X} \right)^{i-n} \right] \ldots \tag{9}$$

Here n represents the number of iterations taken for the blocks for the transaction. Eq. (9) indicates that we can only confirm transactions if we wait for enough blocks. It is important to wait for enough blocks to confirm on blockchains that use PoW or PoS to avoid bifurcation. According to the rules set up on this blockchain, all nodes are allowed to mine. There is no relationship between node count and throughput or verification speed. Consequently, a blockchain network of this type can be scalable to nearly infinity.

3.2 Data Storage

3.2.1 Trustworthy Evaluation Module

Blockchain is a distributed network in which data is shared and stored across a group of IoT nodes. Data collected by IoT sensors can be kept in a distributed blockchain ledger in the context of IoT-driven smart cities. As a result, it removes single points of failure, centralises data, and ensures consistency of the blockchain. Authenticity and actual sensor observations should be represented in the data on the blockchain network. To address the second condition of the architecture at the Device layer, a blockchain reputation system based on IP addresses (Internet Protocol addresses) is developed. Having a high reputation score makes it easier for people to trust the node's observations.

IPFS (Inter Planetary File System) is a global, peer-to-peer, content-addressable, file storage system that has the potential to collect and transmit a large amount of data at high speeds across the Internet. IPFS effectively replaces the traditional technique of sharing documents and data across a network. IPFS is accomplished by joining all devices on the P2P network and unambiguously identifying each file in a global namespace based on content addressing. Each IPFS node is set up to contain a collection of hashed files. As a result, IPFS overcomes the shortcomings of the commonly used client-server approach. The below diagram shows the architecture for trust management and the structure for transactions.

Nodes in a network $Fg = \{fg_1, fg_2, fg_3, \ldots fg_k\}$ is a fog node network that maintains a blockchain ledger for on-chain storage and IPFS for off-chain storage. Peer-to-peer file storage system IPFS allows massive amounts of data to be stored and transported at high speeds thanks to its global, content-addressable distribution. When it comes to storing large amounts of data, blockchain is expensive. However, instead of storing data, it has been demonstrated to be more efficient in storing data in the blockchain. Algorithm 1 describes the stages involved in generating the Reputation Score (RS), which is used to determine the trustworthiness of the blockchain. Each transaction is classified into three categories based on the reputation score. Moreover, IPFS stores raw data in its storage layer according to the requirements and specifications of dishonest and honest data which is then sent to the cloud for long-term storage, while message digests containing reputation score and transaction score of honest, common and dishonest nodes data which are saved in the blockchain network Fig. 4 and Tab. 1 shown the blockchain-based trust management architecture.

Algorithm 1: A reputation calculator based on the address for assessing the reliability of blockchain

1. procedure Compute Reputation (M_{avalue})
2. initialize $M_{ascore} = 0$
   $N_{Threshold} = range(\min(l_j), \max(l_j))$, where $l_j \in D_{set}$
3. Read M_{avalue} of sensor Data is captured from dataset D_{set}
4. if ($M_{avalue} = N_{Threshold}$) then
5. $M_{ascore} += 1$
6. else
7. $M_{ascore} += 0$
8. end if
9. RS = ((M_{avalue})/10/Number of transactions)
10. /* Transaction based on RS and no of features ((f_s) = {f1, f2,…f_n} in dataset*/
11. if (($RS > (f_s.shape - 4)/10$ and ($RS <= (f_s.shape - 4)/10$)) then
12. print("Honest transaction");
13. else if (($RS > (f_s.shape - 8)/10$) and (($RS <= (f_s.shape - 8)/10$)) then
14. print("General Transaction");
15. else
16. print("Dishonest Transactions");
17. end if
18. end procedure

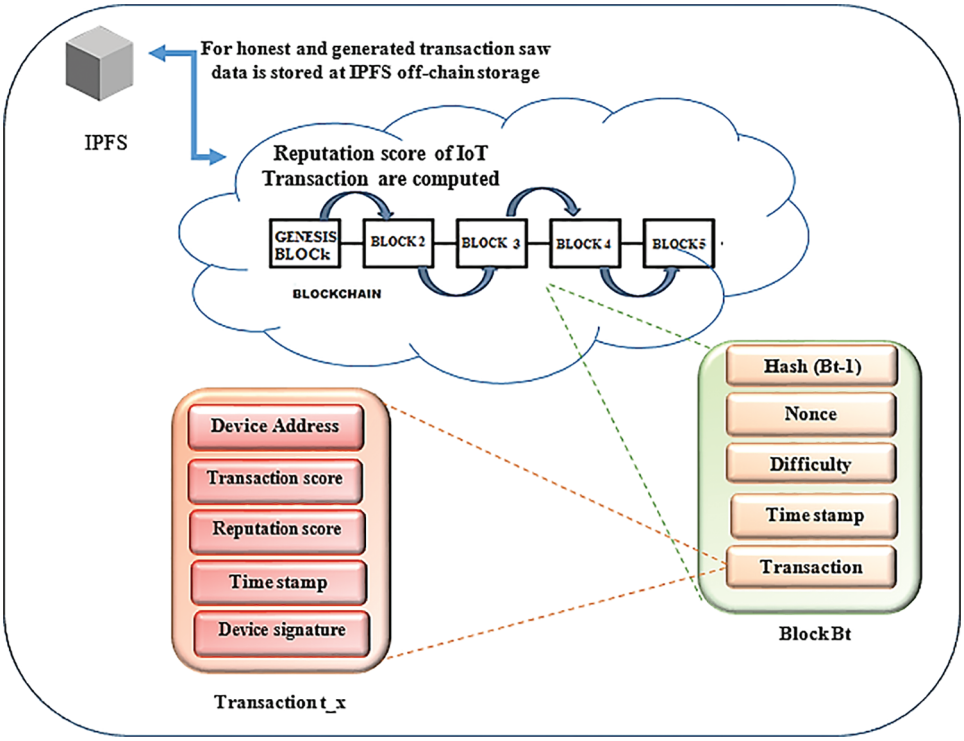

Figure 4: Blockchain-based trust management architecture

Table 1: Comparison table

Time (ms)	BCCA-LSDG	Block BDM	TEMSG	IANFIS
20	70	62	55	35
40	78	72	62	49
60	83	77	71	62
80	91	85	75	69
100	98	92	80	75

3.2.2 Database

The proposed framework is evaluated by applying different experimental datasets that are comprised of graphs.

- AIDS Antiviral Screen Database (48,400 graphs)
- DBLP Dataset (10,000 graphs)
- Graph database (20,000 graphs)

Fig. 5 shows the graphical representation of the graph database. It can choose a significant impact on your company's success. However, consumers frequently have difficulty reconciling the contradictory promises made by various graph software suppliers. The continuous development and growth of huge and sophisticated graph-like data require the use of a graph database. The existence of various graph database solutions demonstrates this renaissance. External interfaces (user interfaces or APIs), database languages (manipulation, querying and data definition), query optimizer, transaction engine, database engine, storage engine management and operation features are all required for the graph databases.

3.3 Service Layer

In this layer, we compute the candidate subgraphs (answer set) that consist of graphs which is similar to q. The initial authentication request from the data user consists of identity, password and secret key which ensure to secure and verify the user. Once the query request arrives from the user to the service provider, it forwards it to the blockchain. The blockchain verifies authentication requests and allows accessing the system.

a) Belief entropy function: A Belief Entropy Function is computed to find the optimum set of features for each graph. This entropy function is based on the Deng Entropy which functions by the probability interval analysis from the lower and upper bound values. From the feature sets for all graphs in the database and entropy values, we retrieve the most optimum set of features.

b) Dual similarity-based Mapreduce: For each feature set, we apply a parallel processing paradigm called MapReduce to choose the subgraph from a large scale database. It is well suited for the real world dataset with decreases in the size and complexity. In the mapping step, subgraphs are determined as structural and semantic similarity.

Fig. 6 shows an example of the similarity between the two graphs. From the two subgraphs G_1 and G_2, the structural and semantic similarity is tested. In structural similarity, the structure of g_1 and g_2 is verified whereas semantic similarity ensures labels of G_1 and G_2.

c) Query log collection: It is one of the novel steps in this research and it is done according to the Query Error Q_e. For each successful query running, we computed the query error. Based on the Q_e, we globally update the index. We use the Post Order Traversal (), which is apart for dynamic updates and avoids the computational overhead and high cost.

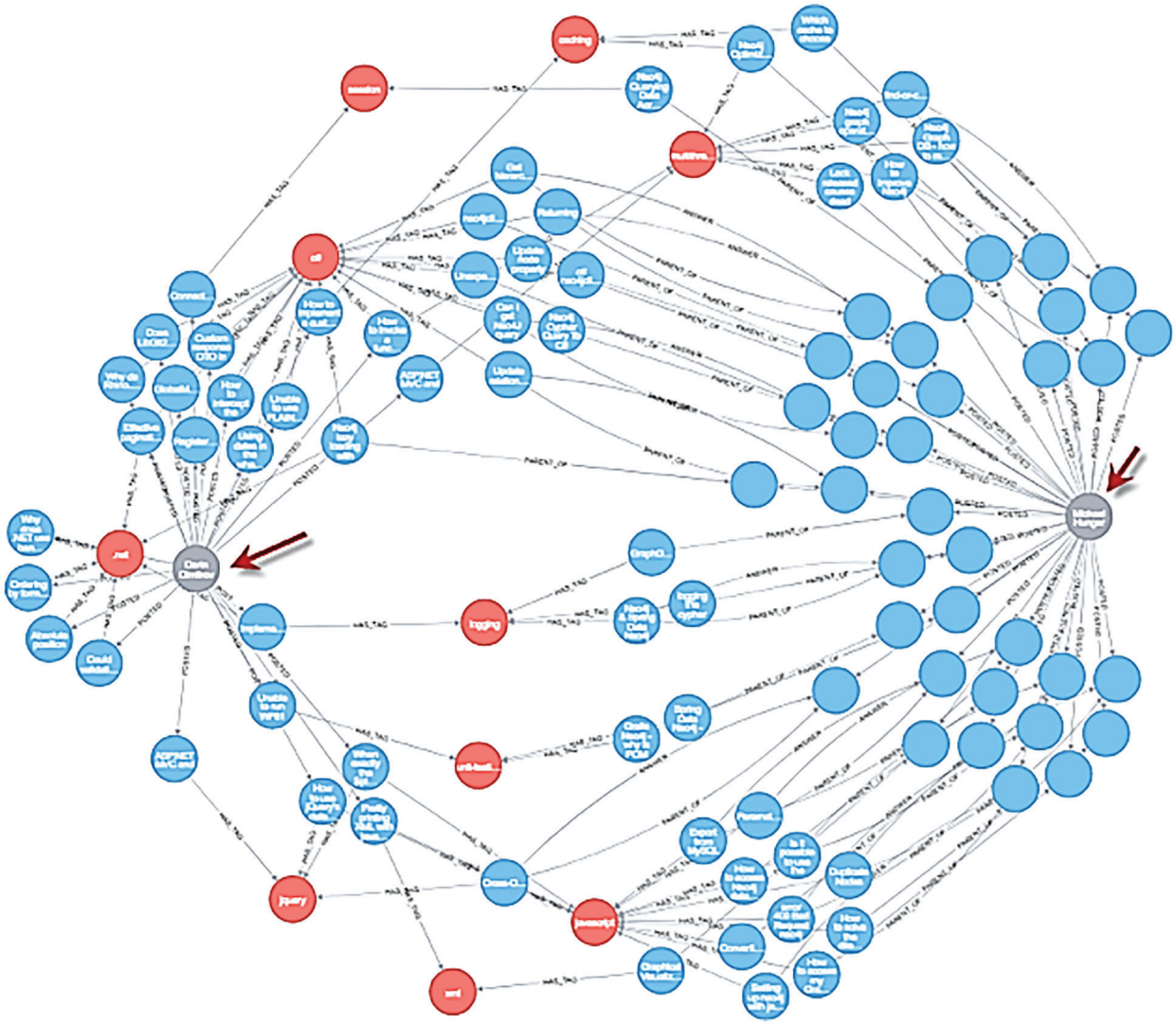

Figure 5: Graphical representation of graph database

Figure 6: An example of the similarity between two graphs

4 Result and Discussion

In this section, we describe an experimental examination of the suggested Blockchain-based consensus algorithm for authenticated query search in the large scale dynamic graphs. The two-fold process is handled

in the proposed BCCA-LSDG. Further, the results are compared with some of the most recent state-of-the-art frameworks in the non-blockchain, blockchain, and hybrid systems.

This section describes about the experiment of the proposed BCCA-LSDG architecture, which is tested on the blockchain enabled Hadoop environment [22]. This architecture is based on Hadoop 2.7.2, Java Development Kit 1.8, Ethereum, and Netbeans 8.0. All the computer software is installed on the Ubuntu 14.04 LTS OS.

Transactions score and Reputation score for 1000 transactions can be seen in Fig. 7. The transactions have been defined as "honest transactions." Various numbers of transactions were investigated with the BCCA-LSDG technique. Transaction score rises steadily with the number of transactions is increased.

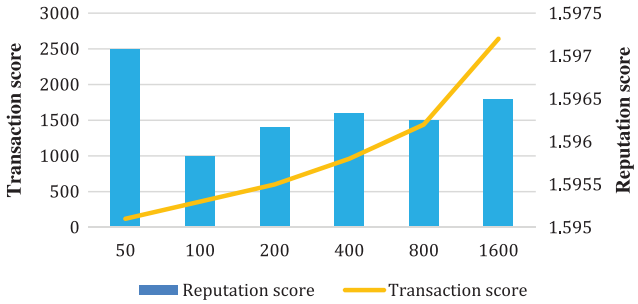

Figure 7: Reputation score *vs.* transaction score for the proposed method

An analysis of hashing power and the likelihood of double-spending success is shown in Fig. 8. Proof of stake (PoS) blockchain performance is limited in order to avoid bifurcations and wait for proper blocks to validate everything [23]. In the PoS ledger, the delegated witnesses construct blocks and verify data transfer.

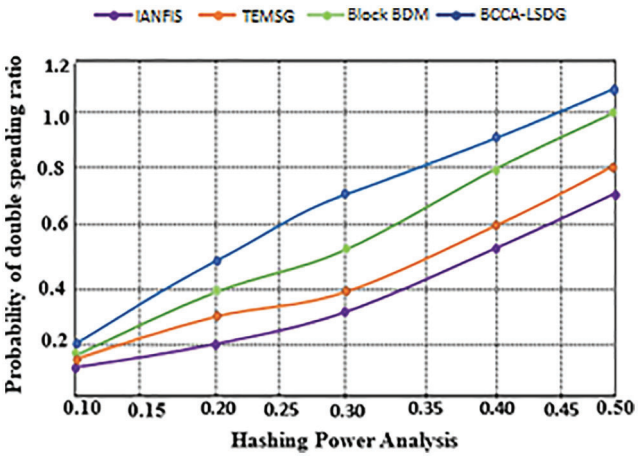

Figure 8: Probability of double-spending and the hashing power

Nodes generate massive amounts of data in a different format across a variety of disciplines, including healthcare, transportation, education, and electricity. There is a lot of information uploaded every day, even if it is possible to see events of the city in real time. Fig. 9 illustrates the importance of having reliable and effective resources for broad data management for the right and successful use of such data in blockchain and the subgraph [24].

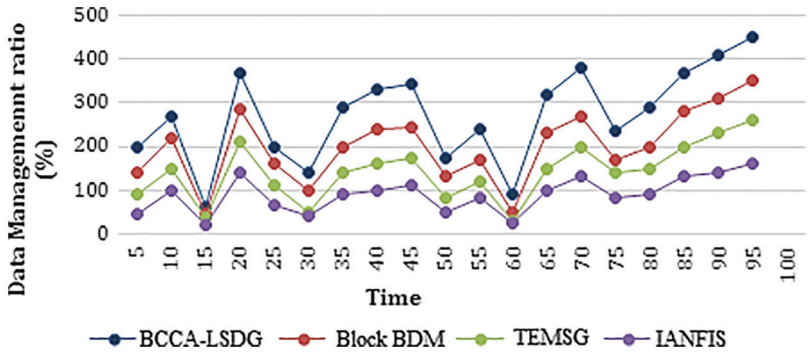

Figure 9: Data management analysis

Fig. 10 demonstrates the comparison of mining accuracy between Block BDM (Blockchain-based Big Data Management) [7], TDMSG (Trust Evaluation Model for Smart Grids) [9], IANFISI (Improved Adaptive neuro-fuzzy inference system) [10], and BCCA-LSDG. According to this definition, mining accuracy denotes the ability of the algorithm to retrieve the relevant outcome for the number of queries entered.

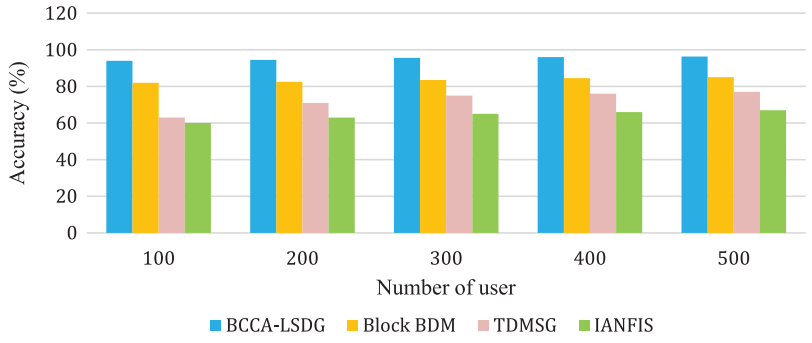

Figure 10: Comparison of accuracy

An individual user's query processing time is measured for the subgraph mining system. According to the processes and algorithms used in mining, this would reflect the variation in the query processing time. A comparison of query processing time with number of users is shown in Fig. 11.

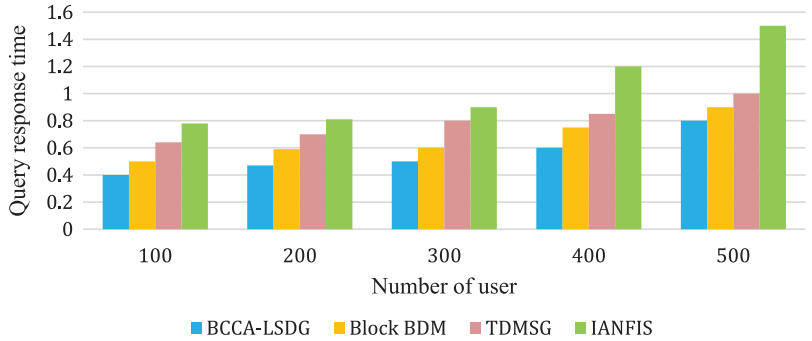

Figure 11: Query response time comparison

Fig. 12 shows the average response time with varying query sizes. Graph Index performs well in all the scenarios and remains consistent with variations in query size. We have verified the response time for varying query sizes as Q4, Q8, Q12, Q16 and Q20 in which the numerals denote the size of the query.

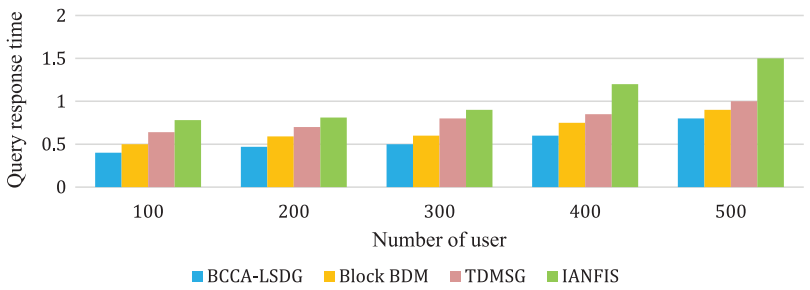

Figure 12: Average response time with varying query size

Fig. 13 shows the performance analysis of proposed technique and the existing methods. The suggested technique uses a PoS consensus algorithm to combine blockchain with a subgraph index. Because blockchain demands a lot of processing resources and takes a long time to finish each operation, the user must make compromises to protect privacy effectively. The proposed BCCA-LSDG scheme gives an optimum solution in terms of overall performance.

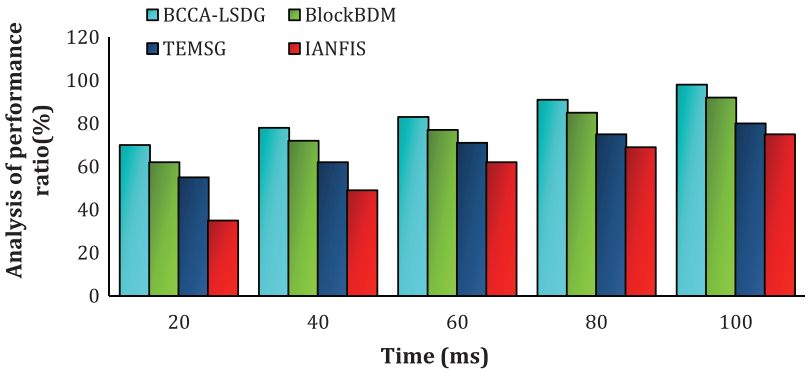

Figure 13: Analysis of the proposed method's performance

5 Conclusion

The aim of the study is to develop a Blockchain-based consensus algorithm for authenticated query search in the large-scale dynamic graphs (BCCA-LSDG) by merging trust evaluation and consensus algorithms to ensure the scalability, security and poor indexing issues in subgraph mining. The proposed consensus algorithm is used for authentication to data owners and users. To overcome the challenges and enable secure huge data transmission, the suggested system incorporates subgraph and blockchain technologies with a trustworthy architecture. Data is kept in the cloud gate server in the blockchain network, and data analysis is handled by the user for ensuring the data security. The accuracy, data management, query response time, and performance of a blockchain are determined by the proposed techniques. Furthermore, the suggested approach offers significant advantages over existing frameworks, both blockchain and non-blockchain. From the results, it can be observed that the overall performance of the proposed method is 98% at 100 ms when compared with existing systems.

Further, three analysis (scalability, storage and security) was conducted to prove the proposed BCCA-LSDG is better than previous works.

Acknowledgement: We would like to thank the supervisors and the anonymous referees for their kind help in this research.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Reference

[1] S. Velampalli and V. R. Jonnalagedda, "Frequent subgraph mining algorithms: Framework, classification, analysis, comparisons," in *Data Engineering and Intelligent Computing*, Singapore: Springer, vol. 542, pp. 327–336, 2018.

[2] Z. Peng, T. Wang, W. Lu, H. Huang, X. Du *et al.,* "Mining frequent subgraphs from a tremendous amount of small graphs using map reduce," *Knowledge and Information Systems*, vol. 56, no. 3, pp. 663–690, 2018.

[3] W. Wang, Y. Yao, L. Zhu, X. Hei, and Y. Wang, "A novel subgraph querying method on directed weighted graphs," in *14th Int. Conf. on Computational Intelligence and Security (CIS)*, Hangzhou, China, pp. 150–154, 2018.

[4] S. U. Rehman, S. Asghar, S. J. Fong, "Optimized and frequent subgraphs: How are they related?" *IEEE Access*, vol. 6, pp. 37237–37249, 2018.

[5] N. T. Le, B. Vo, L. B. Nguyen, H. Fujita and B. Le, "Mining weighted subgraphs in a single large graph," *Information Sciences*, vol. 514, pp. 149–165, 2019.

[6] E. Park, A. P. Del Pobil and S. J. Kwon, "The role of internet of things (IoT) in smart cities: Technology roadmap-oriented approaches," *Sustainability*, vol. 10, no. 5, pp. 1388, 2018.

[7] T. M. Ghazal, M. K. Hasan, M. T. Alshurideh, H. M. Alzoubi, M. Ahmad *et al.,* "IoT for smart cities: Machine learning approaches in smart healthcare—A review," *Future Internet*, vol. 13, no. 8, pp. 218, 2021.

[8] P. Fournier-Viger, C. Cheng, Z. Cheng, J. C. W. Lin and N. Selmaoui-Folcher, "Mining significant trend sequences in dynamic attributed graphs," *Knowledge-Based Systems*, vol. 182, pp. 104797, 2019.

[9] Z. Wang, Y. Zhao, Y. Yuan, G. Wang and L. Chen, "Extreme learning machine for large-scale graph classification based on map reduce," *Neurocomputing*, vol. 261, pp. 106–114, 2017.

[10] T. Renner, J. Müller and O. Kao, "Endolith: A blockchain-based framework to enhance data retention in cloud storages," in *Proc. 26th Euromicro Int. Conf. on Parallel, Distributed and Network-Based Processing (PDP)*, Cambridge, UK, pp. 627–634, 2018.

[11] L. Yue, H. Junqin, Q. Shengzhi and W. Ruijin, "Big data model of security sharing based on blockchain," in *3rd Int. Conf. on Big Data Computing and Communications (BIGCOM)*, Chengdu, China, pp. 117–121, 2017.

[12] W. Dhifli, S. Aridhi, E. M. Nguifo, "MR-SimLab: Scalable subgraph selection with label similarity for big data," *Information Systems*, vol. 69, pp. 155–163, 2017.

[13] M. Zhaofeng, W. Lingyun, W. Xiaochang, W. Zhen, and Z. Weizhe, "Blockchain-enabled decentralized trust management and secure usage control of IoT big data," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4000–4015, 2019.

[14] D. K. Chae, B. K. Kim, S. H. Kim and S. W. Kim, "Incremental feature selection for efficient classification of dynamic graph bags," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 18, pp. e5502, 2020.

[15] B. M. Pampapathi, M. N. Guptha, M. S. Hema, "Data distribution and secure data transmission using IANFIS and MECC in IoT," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 3, pp. 1471–1484, 2021.

[16] K. N. Qureshi and G. Jeon, "A trust evaluation model for secure data aggregation in smart grids infrastructures for smart cities," *Journal of Ambient Intelligence and Smart Environments*, (Preprint), vol. 13, no. 3, pp. 1–18, 2021.

[17] A. S. Yadav and D. S. Kushwaha, "Blockchain-based digitization of land record through trust valuebased consensus algorithm," *Peer-to-Peer Networking and Applications*, vol. 14, no. 6, pp. 3540–3558, 2021.

[18] W. Feng, Y. Li, X. Yang, Z. Yan and L. Chen, "Blockchain-based data transmission control for Tactical Data Link," *Digital Communications and Networks*, vol. 7, no. 3, pp. 285–294, 2021.

[19] L. Cong, J. Yu and X. Ge, "Enabling efficient privacy-preserving subgraph isomorphic query over graphs," *Future Generation Computer Systems*, vol. 132, pp. 1–10, 2022.

[20] Y. Sun, G. Li, J. Du, B. Ning and H. Chen, "A subgraph matching algorithm based on subgraph index for knowledge graph," *Frontiers of Computer Science*, vol. 16, no. 3, pp. 1–18, 2022.

[21] H. Wang, Y. Zhang, L. Qin, W. Wang, W. Zhang *et al.,* "Reinforcement learning based query vertex ordering model for subgraph matching," arXiv preprint arXiv:2201.11251, 2022.

[22] J. Cheng, Y. Ke, W. Ng and A. Lu, "Fg-index: Towards verification-free query processing on graph databases," in *Proc. ACM SIGMOD Int. Conf. on Management of Data*, Beijing, China, pp. 857–872, 2007.

[23] W. Sun, X. Chen, X. R. Zhang, G. Z. Dai, P. S. Chang *et al.,* "A multi-feature learning model with enhanced local attention for vehicle re-identification," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3549–3560, 2021.

[24] W. Sun, G. C. Zhang, X. R. Zhang, X. Zhang and N. N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30803–30816, 2021.