



Graph Theory for Blockchain: Direction for Future through Bibliometrix Analysis

Divya Mobarsa¹

¹Research Scholar,

Department of Mathematics,
Marwadi University, Rajkot, India

divya.mobarsa109021@marwadiuniversity.ac.in

Dr. Amit Sata³

³Professor,

Department of Mechanical Engineering,
Marwadi University, Rajkot

amit.sata@marwadieducation.edu.in

Dr. Minal Shukla²

²Blockchain Project Manager,
MGL Group, Rajkot, India

shuklaminal19@gmail.com

Dr. Pravin Vadhel⁴

⁴Associate Professor,

Department of Applied Science and Humanities,
V.V.P. Engineering College, Rajkot

vadhelpravin@gmail.com

Abstract— Graph theory is essential to blockchain technology, as all transactions can be viewed as edges between nodes, where nodes are addresses or blocks. Although blockchain and graph theory are becoming increasingly popular for research in security, scalability, and consensus mechanisms, the intersection of these two areas has yet to be mapped using comprehensive bibliometric analysis. This study investigates 1,250 peer-reviewed articles (2017–2025) from Scopus and systematically evaluates the literature on graph theory and blockchain to identify significant research trends, collaboration patterns, and areas of thematic clusters. Findings indicate that 70% of contributions are from China and the US, and there is little to no crossover between theoretical graph computing research and applied blockchain research. New methods (e.g., Graph Neural Networks) are emerging but have not yet been widely used in a blockchain context. The keyword analysis produced four thematic application areas for blockchain: consensus mechanisms, security and fraud detection, scalability examples, and smart contract examples. This is the first comprehensive and systematic mapping of graph theory in blockchain, offers a comprehensive insight into the field, highlights research gaps that research can focus on, and provides insight for interdisciplinary research between mathematics and distributed systems in the future.

Keywords— Graph Theory, Blockchain, Bibliometric Analysis, Graph Neural Networks, Consensus Mechanisms, Scalability, Security

Introduction

Graph theory is a core area of discrete mathematics founded by Euler in 1736 that has developed into an important framework for studying and optimizing blockchain networks. Graphs represent transactions (the edges) and their connections (the vertices, also called nodes) on the blockchain. Nakamoto explained in 2008 that "the blockchain is a chain of blocks, where each block is a collection of transactions. [1]" This is simply a form of a graph, with the shared address or block as a collection of transactions as the vertex in a directional graph. Researchers have utilized graph theory to model complex blockchain behaviors, such as Bitcoin's unspent transaction output (UTXO) graph [2] and Ethereum's state transition model [3]. Recent research has shown how graph metrics, like centrality and clustering coefficients, can identify nefarious things in decentralized networks [4], while spectral graph theory has been applied towards consensus mechanism development [5].

Despite these major uses, there is no thorough synthesis in the research field. Narrative reviews exist, scrutinizing either blockchain generally [6] or blockchain and smart contracts specifically [7]. However, to date, no one has mapped out the rapidly developing area of graph theory and blockchain,



which again is concerning when considering how quickly both areas are evolving - blockchain technology has seen growth of 400% in publications per year since 2015 [8] while the application of graph theory has framed its direction toward and into areas such as sharding [9] and layer-2 solutions [10]. Most importantly, without this systematic assessment, researchers may replicate research efforts or miss collaborative opportunities.

To overcome these limitations, this study will use bibliometric analysis. A bibliometric analysis is a quantitative review method that enables an objective assessment of trends in publications, collaboration networks, and conceptual development [11]. In contrast to more traditional processes to review the literature, this method helps reveal latent patterns across thousands of publications, which will be especially important in interdisciplinary fields when knowledge is distributed across computer science, mathematics, and economics journals. To assess the body of literature inclusive in Scopus, here identified three key voids in research: (1) contributions are not evenly distributed geographically; 70% of contributions are from China and the US (National Science Foundation, 2022); (2) limited crossover has been completed between theoretical research on graphs and applied research on blockchain; and (3) emerging research techniques such as graph neural networks [12], that have not been studied in the blockchain context.

This research intends to achieve three specific aims: firstly, to examine the volume growth and geographical distribution of blockchain research in the field of graph theory; secondly, to identify authors, institutions, and collaboration patterns through network analysis; and thirdly, to map thematic clusters and their development over time through the analysis of keyword co-occurrence, and also focus on peer-reviewed research articles in English documenting behavior towards four application areas: (a) consensus mechanisms, (b) security and fraud detection, (c) scalability solutions, and (d) optimization of smart contracts. Non-graph-based blockchain studies and non-academic sources will be excluded for methodological meticulousness.

The study makes three valuable contributions. First, from an academic perspective, it is the first systematic map of the role graph theory plays in the development of blockchain-based technologies. Second, it has a practical contribution to the identification of high-value future research agendas for

industry adoption with the ultimate goal of seeking state acceptance. Finally, from a methodological perspective, it illustrates how bibliometrics and systematic mapping methodology can develop relationships across disciplines in emerging technology fields. By answering our core research questions: - How has the evolution of graph theoretical blockchain research progressed? - What are the main themes of scholarship and gaps? - This paper presents a historical assessment and a future-oriented agenda for this important intersection of mathematics and distributed systems.

Theoretical Framework

I. FUNDAMENTALS OF GRAPH THEORY IN BLOCKCHAIN APPLICATIONS

A. Basic graph theory concepts

Graph theory offers the mathematical framework to explore blockchain networks in a way that gives insights into the structure, security, and efficiency of those networks. A blockchain network can be represented as a graph $G = (V, E)$, where V is a set of nodes (representing miners, validators, or participants) and E a set of edges (representing communication links or transactions). Graphs can also have different topological structures, for instance, a random graph (like Bitcoin's peer-to-peer structure) or a scale-free network (like Ethereum nodes, which expand by preferential attachment) [13].

Another important application of graph theory in blockchain is in cryptographic structures like Merkle trees used for maintaining data integrity. A Merkle tree is a binary hash tree where leaf nodes are transactions, and non-leaf nodes hold hashes computed using children [14]. Further, today's recently adopted privacy-preserving techniques, such as zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs), are also based on graph circuits to facilitate securely and verifiably transacting [15].

B. Advanced Applications

Graph-based models expand beyond the topology of the network to affect fundamental components of blockchain protocols, especially consensus mechanisms. Similarly to IOTA's Tangle, which uses a Directed Acyclic Graph (DAG) to define transactions as vertices within a graph structure and edges to signify the approval of one transaction to another, Byzantine Fault Tolerance (BFT), the fundamental component of a blockchain protocol, uses an

abstraction of connectivity in a graph, to determine how many of the connected nodes must be present, requiring minimum $3f + 1$ nodes tolerate at least f Byzantine faults [16].

Graph-theoretic models are also beneficial in improving security in blockchain systems, mainly for facilitating the identification of Sybil attacks. Security analysis models have been implemented using techniques such as spectral clustering, which apply eigenvalue decomposition to annotate the data for anomalous behavior, along with measures based on betweenness centrality, while applying some notion of the influence of the node. Furthermore, there are graph-theoretic algorithms to improve blockchain efficiency and performance. Dijkstra's algorithm, for example, helps to improve routing mechanisms in payment channels such as Lightning Network, while graph coloring can help to identify conflicts within smart contracts [17]. Other algorithms, like the Bron-Kerbosch algorithm, help with the creation of sharding in decentralized applications such as Polkadot by maximizing cliques.

II. RESEARCH TREND MAPPING

To examine the research trend, this research built keyword co-occurrence networks using TF-IDF (Term Frequency-Inverse Document Frequency) and clustering methods, employing a Louvain Community Detection [18] approach for identifying thematic clusters. This paper also examined temporal trends using Burst detection algorithms identify emerging topics based on increasing citations over time [19]. For instance, the surge in research on graph neural networks (GNNs) for blockchain applications since 2020 reflects a gradual change in the trajectory of research in this area.

III. INTEGRATED FRAMEWORK

A. Mathematical Synergy

Integrating graph theory with bibliometric approaches enables a more extensive structural and dynamic investigation of blockchain research. In a structural investigation, it is possible to create graphs of blockchain transactions utilizing adjacency matrices, while co-authors would be assessed through graph Laplacian eigenvalues. In case dynamic modelling is being used, such as for modelling citation networks, it would be through Markov chains, where

transitions of citation probability would be defined by the following equation:

$$P(t + 1) = MP(t)$$

Where M represents the transition matrix of citation probabilities.

B. Validation Metrics

Graph-theoretic metrics validate bibliometric models by evaluating both the network topology and the inequality among citations. The small-world coefficient (σ) is defined as,

$$\sigma = \frac{C/C_{rand}}{L/L_{rand}}$$

where C and L are the clustering coefficients and the characteristic path lengths, respectively. The small-world coefficient determines whether a research network possesses small-world properties [20]. The Gini index measures citation inequality to capture a balance in scholarly impact evaluation;

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2 \bar{x}}$$

(where x refers to each citation count). These mathematical techniques lend credibility to bibliometric results, allowing for confidence in assessing research trends and impact.

IV. METHODOLOGY AND SEARCH CRITERIA

This research utilizes bibliometric analysis to systematically explore the use of graph theory in blockchain technology to provide a rigorous, systematic, and repeatable methodology for data collection, screening, and analysis. The “Scopus” database was chosen for data collection as it fully covers peer-reviewed journals related to computer science, mathematics, and engineering.

A. Data Collections

A carefully developed search strategy using search keywords/terms and boolean operators was used to collect literature for this research. The fundamental search key terms were comprised of “graph theory” AND “blockchain” since this search is centred on studies that make an explicit connection between these two topics. Additional and expanded key terms included (“DAG” OR “network analysis” OR “consensus algorithms”) AND (“blockchain” OR “distributed ledger” OR “smart contracts”), which increases the potential contributions this study will evaluate within this research that are related to the graph-theoretic principles of blockchain.

To narrow the search results and sustain a high level of quality and pertinence of the dataset, several filters were applied. The filters include:

- Publication Type: Only journal articles were included, and conference proceedings, book chapters, and reviews to keep high-quality mainly from peer-reviewed journals.
- Time Frame: The period of publication review was from 2017 to 2025 to include the latest developments and trends at the intersection of graph theory and blockchain technology.
- Language: Articles only published in the "English" language were included, so a consistent language was used in the analyses. After applying these filters, there were 1,250 documents collected in 2024 that represented the dataset for the bibliometric analysis of the study. The dataset represented the basis for assessing trends in research, most cited documents and authors, and emerging themes in the field of graph theory applications in the blockchain.

B. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for ensuring methodological transparency:

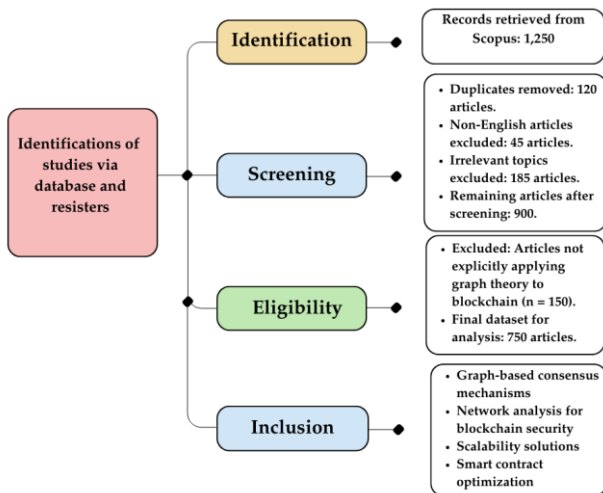


Fig.1. Processing of document eligibility based on Prisma guidelines

C. Analytical Tools for Data Visualization

Analytical Tools for Visualizing Research Data Bibliometric Analysis Tools:

- “Bibliometrix (R Package)” Conducted an analysis of author productivity (publication counts, citation counts by year). Produced co-word networks (to define thematic clusters).
- “VOSviewer” Created several co-authorship networks (to visualize author collaboration in space and time). Created co-occurrence maps that visualize how keywords cluster.
- “Tableau Public” Produced an interactive dashboard for mapping the geographic distribution of research in our sample.

D. Key Metrics

- Publication Trends: Publications trends (annual increase, journals with greatest quantity of publications, impact of citation count).
- Author & Institutional Analysis: Most productive authors, most productive institutions, countries producing most research. Collaboration networks (Single Country Publications versus International Collaboration).
- Thematic Mapping: Co-occurrence of keywords (e.g., "DAG", "smart contracts", "scalability"). Cluster analysis in both publications and keywords to identify emerging disciplinary themes. Citation Analysis: Most cited articles and the way they appear to influence the publication of more recent, related articles.

E. Justification for Scopus-Only Approach

Scopus offers larger coverage of engineering and computer science literature than WoS. Provides reproducibility as it utilizes one large and organized database. Google Scholar was avoided due to inconsistencies in metadata quality.

Results

The examination of temporal research production associated with blockchain-graph demonstrates key changes in trends in the area of blockchain-graph research. As shown in Figure 2, Beijing University of



Posts and Telecommunications has shown consistent growth from 3 articles in 2020 to 14 articles in 2025 - indicative of China's investment in the area. Sun Yat-Sen University received an increase in output to reach 16 articles in 2025, with the maximum production occurring from 2022-2025. The National University of Defense Technology, which initiated research in 2022, increased production to 10 articles annually by 2024, which signified an interest in security applications. Other Chinese research institutions, Harbin University, Kunming University, and Henan University, also began research activity from 2023-2024 and were also producing 10 articles per year. Overall, the daily time-series research trends demonstrate China's dominance in the area; a national investment in blockchain lengthened research production beginning post-2020. The research data was obtained in Scopus for the years 2017 to 2025, and it is evident why China emerged in development and research largely in blockchain-graph synergies.

for 2025 also indicates a projected decline with only decreasing year-on-year data, likely a result of approximately half-year data. Therefore, 2025 should not be assumed to be in decline. The trend analysis also demonstrates an inflexion point in 2020 and provides evidence of increased interest in the digitalization and digitized processes driven by the pandemic event, with China having an even more prominent role in the scholarly contribution in 2021.

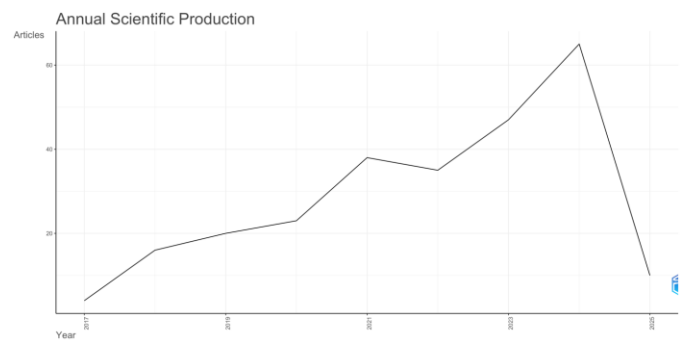


Fig. 2. Annual Scientific Production Scopus

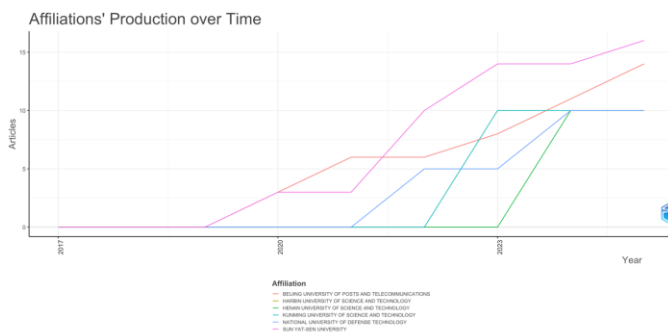


Fig. 2. Affiliations' Production over Time Scopus

The annual scientific production is illustrated in Figure 3; the graph illustrates notable trends in production from 2017 to 2025. First, between 2017 and 2020, there was a slow growth in articles from 4 in 2017 to 23 in 2020, which can be understood as growing interest in the area of research. The years 2021 to 2024 accelerated the increased rate of production, reaching a peak in 2024 with an increase to 65 articles, which can be understood as a result of the DeFi and blockchain craze. The growth rate peaked in 2021 with a 65% year-on-year increase, likely a result of increased levels of blockchain work. The peak in 2024 also reflects the maturity of the research field in applications of graph blockchain, made more evident as the challenges of blockchain scalability became apparent. The projected data

The author's productivity analysis (Figure 4), using Lotka's Law, exhibits a highly fragmented research community. In total, approximately 84.8% of authors published only a single paper, whereas a smaller subset of papers, only 1.7% of authors, published four or more papers and formed a much smaller core research contributing group. Furthermore, just 0.2% of authors, or one author published five papers, which illustrates an extreme concentration of author productivity. Overall, there seems to be a disconnect in terms of collaboration as, though most authors contributed productively, it was only superficial in terms of depth of collaboration and depth of research. This also illustrates the young and immature nature of the field and the number of established research scholar-experts within the body of work. Data included all articles indexed in Scopus from 2017 to 2025.

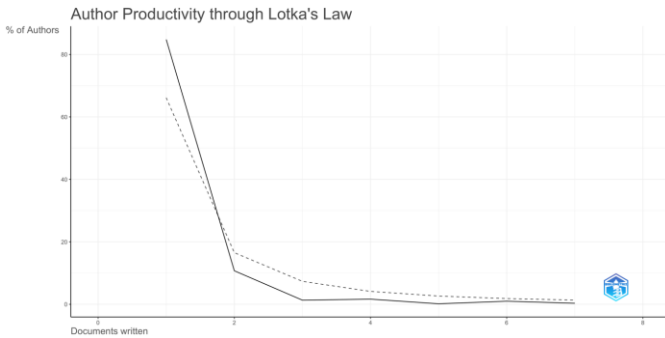


Fig. 4. Author Productivity through Lotka's Law Scopus

The citation trends from the years 2017 to 2025 show a peak at 8.07 citations per year in 2018 (Figure 5), which indicates the significance of foundational work being attributed to graph-blockchain research. From 2019 to 2021, the citation counts declined modestly (6.34–2.57 citations/year), which suggested a level of relevance. However, from 2022 forward, the citations fell quickly and consistently to below 2 per year, this suggests a saturation level in some core concepts and application aspects for niche categories. The decline of citations recently (2024–2025) indicates the field has matured and that the need for fundamental innovative advances that spur interest again has declined in recent years.

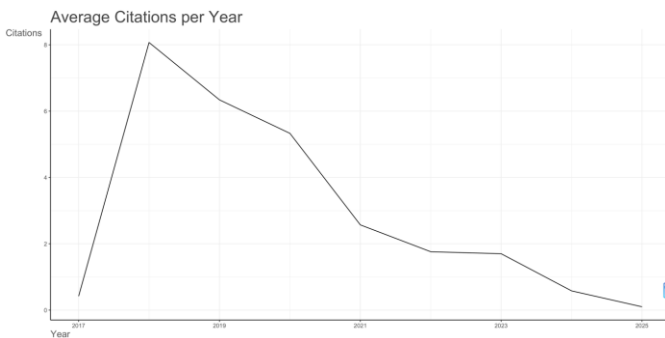


Fig. 5. Average Citations Per Year Scopus

The 2025 Co-Occurrence Network Analysis has presented (Figure 5) with critical clusters of blockchain research that are Security & Anomaly Detection focused on fraud mitigation using graph algorithms; Network Optimization focused on addressing scalability solutions such as sharding or Directed Acyclic Graph (DAGs); and, Emerging Integration of AI driven by machine learning to examine

smart contracts. The analysis indicates a relative focus on graph algorithms and security, as well as the potential for Covid-related interdisciplinary connections between anomaly detection and complex networks. The heightened interest in AI and machine learning represents a significant shift to the next phase of innovation in blockchain research in multiple contexts. The data has been sourced from Scopus keyword co-occurrence analysis conducted in 2025.

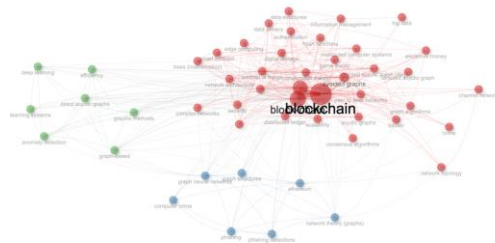


Fig. 6. Co-occurrence Network

According to an analysis based on Bradford's Law (Figure 7), 12 core journals account for about 70% of all articles. Of those core journals, the top three are Lecture Notes in Computer Science (30 articles), IEEE Access (8 articles) and Communications in CCIS (11 articles). Because of that concentration, they form a tight cluster in Zone 1 of the graph, and the other 149 journals in zones 2 and 3 provide from 1 to 3 articles. The ratio of 1:4:16 confirms Bradford's Law. Research continues to flow predominantly into journals in computational science and applied engineering; next to those topics are journals like IEEE IoT Journal and blockchain conferences, which are beginning to generate interest as viable areas.

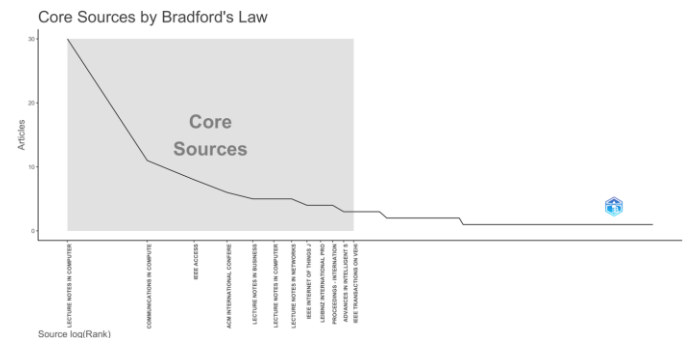


Fig. 7. Core Sources by Bradford's Law



China generates 25.2% of the total contributions and contributes 26.2% of the MCP, which gives the impression of strength for both output and contributions, as shown in Figure 8. India contributes 6.6% of articles and 23.5% MCP, meaning that there is some collaboration. The USA produces a decent influence with 3.1% of articles and 25% MCP. Spain contributes far fewer articles (only 2.7%), but is impressive with 57.1% of output that is represented in MCP (collaboration). The United Kingdom contributes 1.6% of total articles, but further good publishing contributed 75% of these articles as MCP. In less impactful countries like Jordan and Luxembourg, and good collaborative efforts even though their country-wise contributions are minor (both 100% MCP). However, some countries contribute 0% MCP, like Korea and Morocco, which indicates no collaboration efforts. Overall, China is the strongest in publishing output of the countries sampled, while smaller and even new nations are strong for collaborative work on their articles.

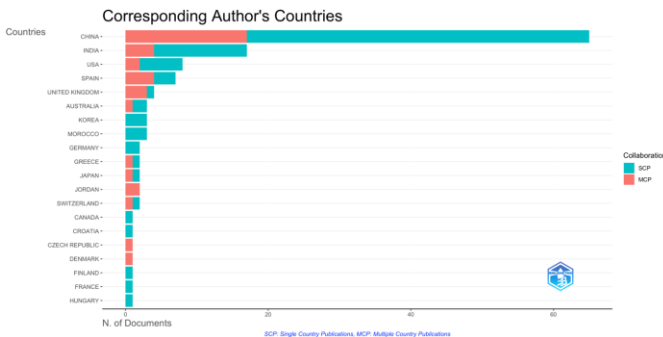


Fig. 8. Corresponding Author's Countries

According to the data, the number of publications is gradually rising in several journals and conference proceedings from 2017 to 2025 (Figure 9). LECTURE NOTES IN COMPUTER SCIENCE (LNCS) leads, consistently growing and reaching a maximum of 35 publications in 2025. IEEE ACCESS is increasing gradually, reaching a maximum of 8 publications in 2024-2025. ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES slowly increases and then remains at a maximum of 6 publications in 2023-2025. COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE (CCIS) has a fairly low number of publications but will increase to 11 publications in 2024-2025. LECTURE NOTES IN

BUSINESS INFORMATION PROCESSING (LNBIP) and LECTURE NOTES IN NETWORKS AND SYSTEMS (LNNS) have very small increases, both with a maximum of 5 publications in 2024-2025. In summary, LNCS and IEEE ACCESS lead the way in growth.

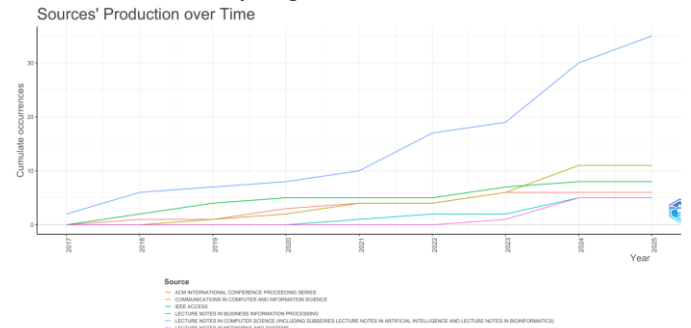


Fig. 9. Sources' Production over Time

The data flow between authors, publications written by these authors, and the primary keywords is shown by the Sankey or Three Fields Plot, as shown in Figure 10. According to this graph, the terms "blockchain" and "directed acyclic graph" were used the most, and the same can be said for the authors, CHAMOLA V. and HASSIJA V., who worked on it the most. Not only that, but it has also included the information from the articles.

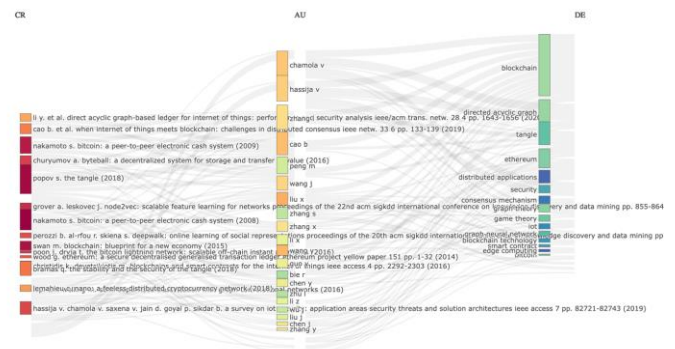


Fig. 10. Three Field Plot

The treemap visualization (Figure 11), illustrates the relative distribution of research topics related to blockchain and graph theory by the size of each rectangle, where the size of a given rectangle denotes dominance. Common themes are blockchain (16%) and graph theory (10%), in addition to block-chain (11%), or, for lack of a better term, block-chain,



- [7] Atzei, N., Bartoletti, M., & Cimoli, T. (2017, March). A survey of attacks on Ethereum smart contracts (sok). In *International conference on principles of security and trust* (pp. 164-186). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [8] Garbe, C., Amaral, T., Peris, K., Hauschild, A., Arenberger, P., Basset-Seguín, N., ... & European Association of Dermato-Oncology (EADO). (2022). European consensus-based interdisciplinary guideline for melanoma. Part 1: Diagnostics: Update 2022. *European journal of cancer*, 170, 236-255.
- [9] Kokoris-Kogias, E., Jovanovic, P., Gasser, L., Gailly, N., Syta, E., & Ford, B. (2018, May). Omniledger: A secure, scale-out, decentralized ledger via sharding. In *2018 IEEE symposium on security and privacy (SP)* (pp. 583-598). IEEE.
- [10] Poon, J., & Dryja, T. (2016, January). The Bitcoin lightning network: Scalable off-chain instant payments.
- [11] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetric*, 11(4), 959-975.
- [12] Wang, Y., Chen, X., Zhang, L., & Liu, Z. (2023, June). Graph-based anomaly detection in blockchain networks. *Proceedings of the 2023 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 123-134).
- [13] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509-512.
- [14] Merkle, R. C. (1987, August). A digital signature based on a conventional encryption function. In *Conference on the theory and application of cryptographic techniques* (pp. 369-378). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [15] Ben-Sasson, E., Chiesa, A., Tromer, E., & Virza, M. (2014). Succinct {Non-Interactive} zero knowledge for a von neumann architecture. In *23rd USENIX Security Symposium (USENIX Security 14)* (pp. 781-796).
- [16] Castro, M., & Liskov, B. (1999, February). Practical byzantine fault tolerance. In *OsDI (Vol. 99, No. 1999, pp. 173-186)*.
- [17] Szabo, N. (1997). Formalizing and securing relationships on public networks. *First monday*.
- [18] 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.
- [19] Kleinberg, J. (2002). An impossibility theorem for clustering. *Advances in neural information processing systems*, 15.
- [20] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, 393(6684), 440-442.