南京大學

FINAL REPORT

Topic: Blockchain Publications in Web of Science from 2020 To 2024

Subject: The Organization of Information

Professor: Jiaqi Yan, Prof.

Student Name: Do Minh Ngoc

Student ID: 502024145013

Table of Contents

Table of Figures

List of Tables

Abstract

In the context of increasing number of scientific publications on blockchain technology, organizing, standardizing and analyzing academic data networks play an important role in understanding the knowledge structure, research trends and collaboration patterns in this field. This study proposes an integrated process from raw data processing to network analysis to explore the co-authorship and citation relationships in blockchain-related SCI papers. The initial data is organized, standardized and visualized using Excel; then it is transformed into a relational database model and academic ontology using Protégé, with entities such as Author, Article, Subject, Journal and Citation.

Based on the processed data, co-authorship and citation networks are constructed and analyzed using Gephi. Algorithms such as Louvain (Modularity) and network metrics (Degree Centrality, Betweenness Centrality, Density, Diameter...) are used to detect research communities, identify central authors, and evaluate the structure of scientific collaboration. In addition, the time-series analysis (2020–2024) shows the transition from a fragmented network to a recent trend of stronger collaboration convergence. The results contribute to providing a quantitative and intuitive view of the development dynamics of blockchain research from the perspective of academic data and knowledge networks.

1. Introduction

1.1. Reasons for choosing the topic

The project was carried out based on the course requirements, with the aim of synthesizing the knowledge learned about data organization, database design, ontology construction and network analysis in a real-world context. Instead of choosing hypothetical data, the group decided to exploit a collection of scientific articles related to the topic of blockchain - a rapidly developing field with a large number of publications. Building and analyzing the network of co-authors, citations and topics from academic data not only meets the course requirements, but also provides a practical perspective on the collaboration structure and research trends in this emerging field.

1.2. Research question

In this paper, we analyze the trends and structural efficiency of the co-authorship network within the blockchain research field—an emerging and rapidly growing discipline in the domain of technology.

1. **How are citation relationships formed among blockchain-related scientific publications?**
2. **Which articles or authors occupy central positions in the literature network in terms of citation and collaboration?**
3. **Are there distinct research communities identifiable through citation and co-authorship patterns?**

1.3. Research object and scope

The research object of the topic is a collection of scientific articles in the SCI category with content related to blockchain technology, including information about authors, organizations, journals, citations and research topics. On that basis, the thesis focuses on organizing, processing and analyzing this academic data from many aspects: organizing data tables in Excel, designing relational databases, building academic ontology and analyzing co-author networks using the Gephi tool.

The scope of the research is limited to articles published in the period from 2020 to 2024, with the aim of exploring development trends, academic collaboration structures and the formation of research communities in the field of blockchain.

1.4. General research methods

The topic applies an empirical research method based on secondary data. The research process includes the following main steps:

1. Organizing and standardizing raw data using Microsoft Excel.
2. Designing a relational database to separate article information, authors, and citation relationships.
3. Designing an academic ontology using Protégé software.
4. Building and analyzing the author collaboration network using Gephi. Network analysis techniques, such as community detection, centrality calculations, density, diameter, and temporal analysis, are employed.

1.5. Thesis structure

The report is organized into six chapters as follows:

- **Chapter 1**: Introduction
- **Chapter 2**: Theoretical Basis and Research Overview
- **Chapter 3**: Organizing and Preprocessing Academic Data (Excel) and Database Design (SQL)
- **Chapter 4**: Ontology Model (Protégé)
- **Chapter 5**: Analyzing Co-authorship and Citation Networks (Gephi)
- **Chapter 6**: Conclusion.

2. Theoretical basis and research overview

2.1. Theoretical basis of academic collaboration network

Co-authorship in research is a clear indicator of academic collaboration and can be modeled as a collaboration network where each author is a node, and the co-authorship relationship is

represented as an edge. Social Network Analysis (SNA) techniques are used to evaluate the role, centrality, and connections of each individual within the scientific community. Key concepts in this domain include:

- **Degree Centrality**: Measures the level of connection an author has with other authors.
- **Betweenness Centrality**: Identifies authors who act as intermediaries between different research groups.
- **Modularity**: Used to detect communities or groups of authors who frequently collaborate.
- **Small-world Network**: A typical feature of academic networks, indicating high connectivity and short average distances between nodes.

2.2. Research related to collaboration networks in the Blockchain field

Recent studies have begun to apply network analysis to the blockchain field. For instance, Hassanein et al. (2025) analyzed the co-authorship network in the blockchain accounting domain and found that collaboration concentrated around a few key authors, with limited international collaboration in developing countries. Similarly, Kim (2024) examined the collaboration network within blockchain consensus mechanisms, identifying large collaborative communities and central figures such as Aggelos Kiayias. These studies confirm the utility of network analysis in exploring the development and social structure of blockchain research.

2.3. Research Gap and Motivation for the Project

Although there have been some studies on co-authorship networks in the Blockchain field, most of them still focus on basic statistics (number of articles, citations, keywords) without closely combining with systematic data organization and ontology modeling. In addition, there is a lack of time-based analysis to observe the development of the collaboration network over the years.

This project aims to build an overall process - from processing raw data, standardizing information using data models and ontology, to mining collaboration networks using Gephi - to contribute to filling this gap in Blockchain research.

3. Organizing and pre-processing academic data

3.1. Organizing initial data using Excel

3.1.1. Structure of "Raw Data" table

The research data was collected from the **Web of Science (WoS)** database, filtered by articles related to the topic of **Blockchain** published in journals indexed in the **Science Citation Index (SCI)**. This dataset is stored in a worksheet named **"Raw Data"** and is kept unchanged to preserve its original integrity.

The data table includes the following main information fields:

Table 1. Structure of Raw Data

Field Name	**Description**
UT	Unique identifier for each article in the Web of Science system
Author_Full_Names	Full names of all authors (formatted according to WoS standards)
Article_Title	Title of the article
Source_Title	Name of the journal where the article was published
Author_Keywords	Keywords provided by the authors
Keywords_Plus	Additional keywords suggested by the system based on citation terms
Abstract	Abstract of the article
Addresses	Institutional affiliations and addresses of the authors (e.g., university, country)
Publication_Year	Year of publication
DOI	Digital Object Identifier of the article
WoS_Categories	Subject categories assigned by Web of Science
Research_Areas	Primary research areas of the article

3.1.2. Processed Data

Following the data cleaning and structuring process in Excel, the raw dataset was transformed into a normalized format in the "Processed Data" worksheet. Each author of a multi-authored article was assigned a separate row, accompanied by an "Author Order Number" column to preserve authorship hierarchy. Similarly, institutional affiliations were expanded so that each address was tied explicitly to its respective author.

The dataset revealed that the maximum number of authors in a single publication was **28**, found in the article titled *"Connecting supplier and DoD blockchains for transparent part tracking."* Meanwhile, the article *"Unchaining Collective Intelligence for Science, Research, and Technology Development by Blockchain-Boosted Community Participation"* had the highest number of unique institutional affiliations, with **13 distinct author addresses**. Kumar, Neeraj has 44 addresses — the highest number among all authors.

3.1.3. Processing and Theme Selection Rationale

Categorization

To facilitate coherent analysis and effective visualization in the subsequent steps (e.g., pivot tables and charts), a main theme was selected for each paper. Specifically, **only the first listed WoS category** was retained, with the following rationale:

- **Avoiding data fragmentation and duplication**: Articles often belong to multiple overlapping subject categories (e.g., *"Agricultural Economics & Policy; Economics"*). Including all would cause the pivot table to treat each combination as a separate entity, leading to fragmented groupings and less meaningful aggregation.
- **Representing the primary research domain**: In Web of Science and similar databases, the first category listed typically reflects the article's core disciplinary focus, whereas subsequent categories are often peripheral. Thus, using only the first theme ensures clarity in classification.
- **Enhancing chart readability**: By limiting each paper to a single primary theme, resulting pivot tables and bar charts are simpler, with fewer categories, making visual comparisons across years or topics more interpretable.

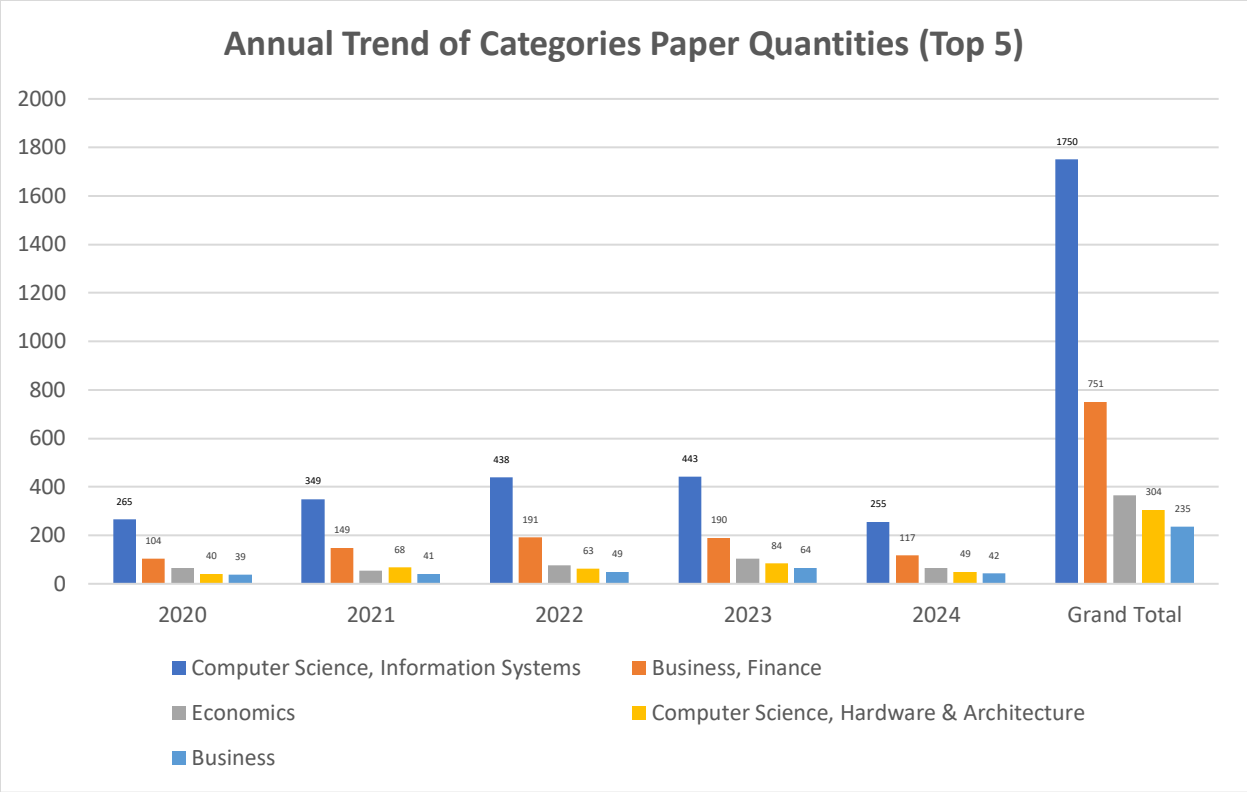

Figure 1. Annual Trend of Categories Paper Quantities (Top 5)

From 2020 to 2024, Computer Science, Information Systems consistently dominated blockchain research, followed by Business, Finance and Economics, reflecting the technology's core and its financial applications. In contrast, fields like Quantum Science, Sociology, and Medical Imaging saw almost no blockchain-related publications, indicating limited interdisciplinary integration in these areas so far.

Additional Observations from Descriptive Statistics

Key word Analysis

- **Core Concepts**: Keywords such as *blockchain*, *cryptocurrency*, *smart contract*, and *bitcoin* appeared frequently across the dataset, confirming their foundational role in this research domain. *Ethereum*, the most widely used platform for decentralized applications (DApps) and decentralized finance (DeFi), consistently ranked among the top keywords due to its popularity as a case study platform.

- **Topical Trends**: In the period 2021–2022, keywords such as *COVID-19* surged in frequency, reflecting a wave of studies exploring blockchain applications in pandemic-related contexts, especially in healthcare and supply chain transparency.

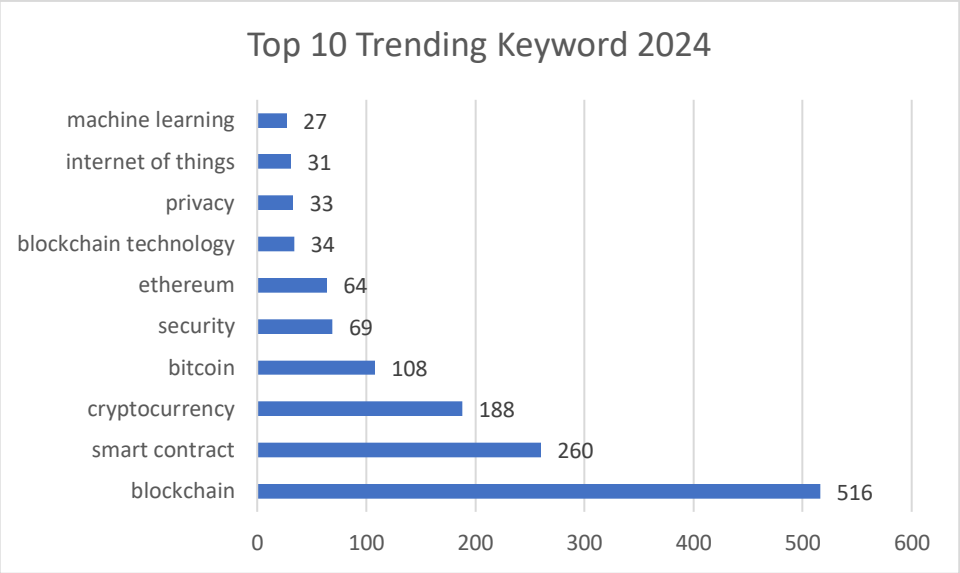

Figure 2. Top 10 Trending Keyword 2024

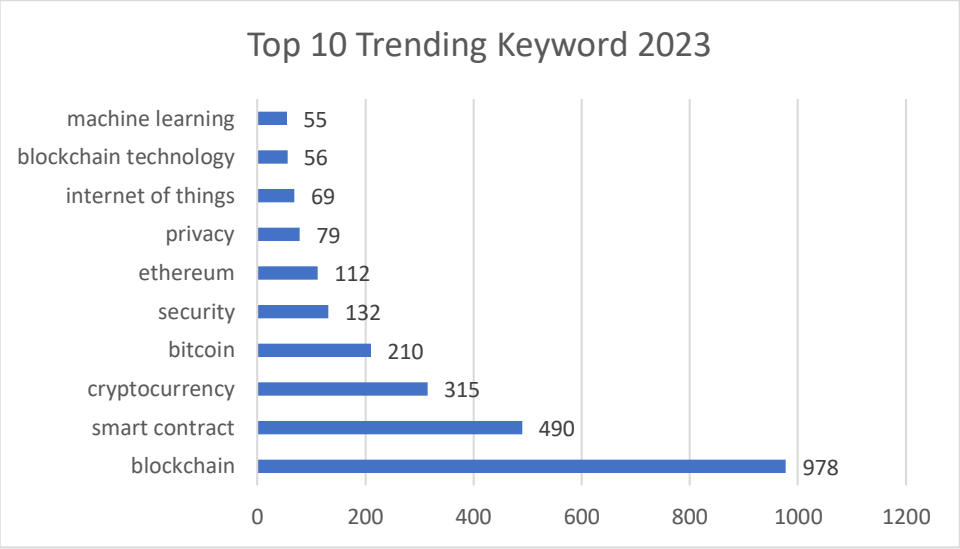

Figure 3. Top 10 Trending Keyword 2023

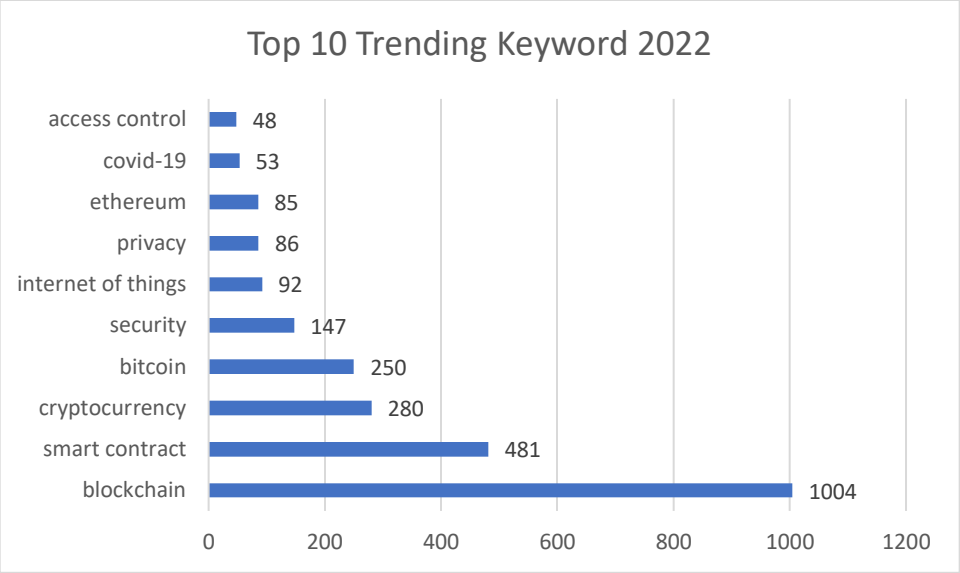

Figure 4. Top 10 Trending Keyword 2022

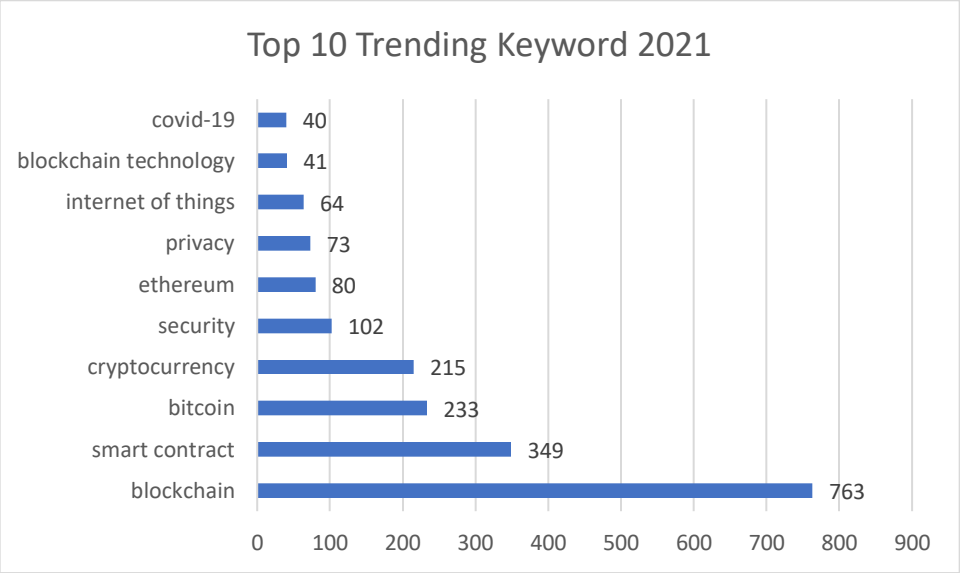

Figure 5. Top 10 Trending Keyword 2021

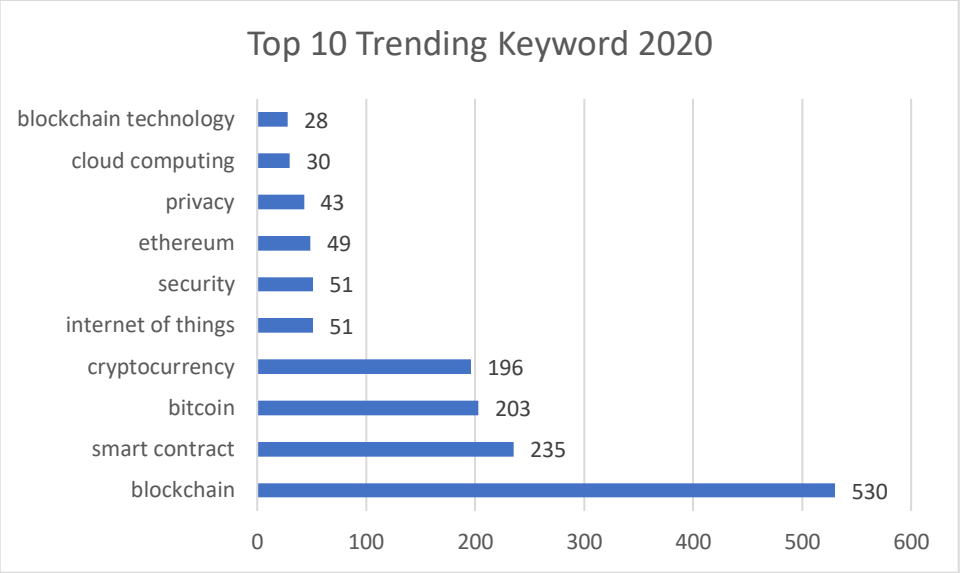

Figure 6. Top 10 Trending Keyword 2020

- **Geographical Trends**: The majority of authors originated from **China** and the **United States**, which aligns with both countries' status as global technology hubs. Their leadership in blockchain R&D—supported by robust academic, governmental, and industrial ecosystems—translates into high publication output. This trend underscores the influence of national-level policies and market interests on academic research volume and focus.

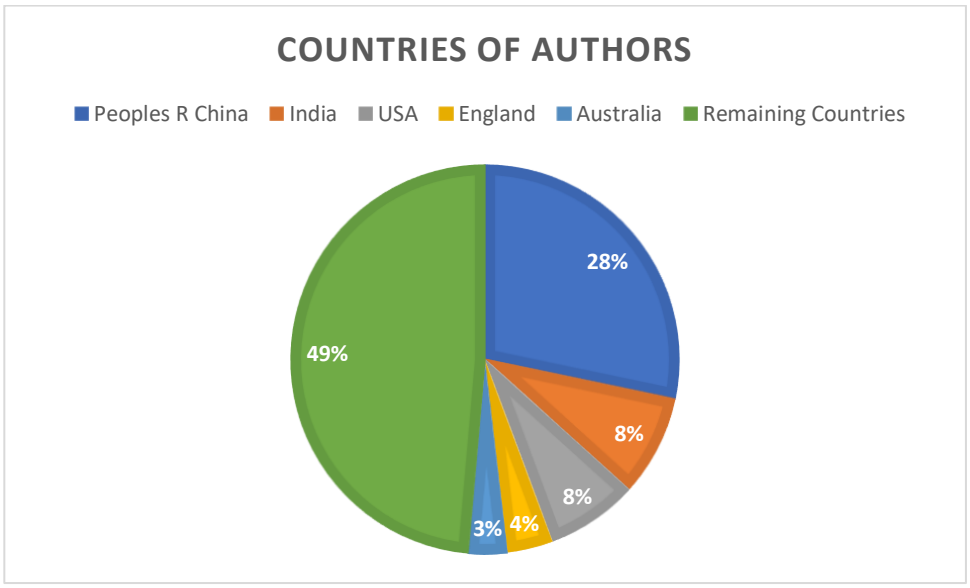

Figure 7. Bar Chart revealed Countries of Authors

3.2. Designing relational database (MySQL/SQL Server)

3.2.1. Creating table and importing data

The process of designing a relational database for organizing academic data related to blockchain research involves creating structured tables, importing the data, and establishing relationships between entities like **Papers**, **Authors**, **Institutions**, and **Citations**. The database is designed using SQL, which provides the foundation for querying and analyzing the data.

SQL Querry:

Paper_Basic_Info_Table:

This table holds basic information about each paper, including the title, abstract, publication year, journal name, DOI, categories, and keywords.

```
CREATE TABLE Paper_Basic_Info_Table (
    Paper_ID VARCHAR(255) PRIMARY KEY,  -- UT (Unique identifier)
    Article_Title TEXT,
    Abstract TEXT,
    Publication_Year INT,
    Journal_Name TEXT,                  -- From Source_Title
    DOI VARCHAR(255),
    WoS_Categories TEXT,
    Research_Areas TEXT,
    Author_Keywords TEXT,
    Keywords_Plus TEXT
);
```

Author_Info_Table: This table contains details about the authors, such as their names and affiliated institutions.

```
CREATE TABLE Author_Info_Table (
    Author_Name VARCHAR(255) PRIMARY KEY,
    Affliated_institution TEXT
);
```

Author_Address_Table: This table stores addresses associated with each author. If an author has multiple addresses, each address is recorded on a new row, with a foreign key relationship to the **Author_Info_Table**.

```
CREATE TABLE Author_Address_Table (
    Author_Name VARCHAR(255),
    Address_Order INT,
    Address TEXT,
    PRIMARY KEY (Author_Name, Address_Order),
    FOREIGN KEY (Author_Name) REFERENCES Author_Info_Table(Author_Name)
);
```

Paper_Author_Relation_Table: This table represents the many-to-many relationship between **Papers** and **Authors**. It connects each paper with its respective authors and records the order of authorship.

```
CREATE TABLE Paper_Author_Relation_Table (
    Paper_ID VARCHAR(50),
    Author_Name VARCHAR(255),
    Author_Order INT,
    PRIMARY KEY (Paper_ID, Author_Name),
    FOREIGN KEY (Paper_ID) REFERENCES Paper_Basic_Info_Table(Paper_ID),
    FOREIGN KEY (Author_Name) REFERENCES Author_Info_Table(Author_Name)
);
```

Citation_Relation_Table: This table records citation relationships between papers, including the citation count. The **Citing_Paper_ID** and **Cited_Paper_ID** columns represent the relationship between two papers, and the citation count is captured for analysis.

```
CREATE TABLE Citation_Relation_Table AS
SELECT
    `from` AS Citing_Paper_ID,
    `to` AS Cited_Paper_ID,
    COUNT(*) AS Citation_Count
FROM
    cite_relationship
```

```
GROUP BY
  `from`, `to`;

ALTER TABLE Citation_Relation_Table
MODIFY Citing_Paper_ID VARCHAR(50),
MODIFY Cited_Paper_ID VARCHAR(50),
ADD PRIMARY KEY (Citing_Paper_ID, Cited_Paper_ID),
ADD FOREIGN KEY (Citing_Paper_ID) REFERENCES Paper_Basic_Info_Table(Paper_ID),
ADD FOREIGN KEY (Cited_Paper_ID) REFERENCES Paper_Basic_Info_Table(Paper_ID);
```

With the database populated, SQL queries are executed to retrieve key insights from the data. These queries help identify trends in research publications, such as the top journals by paper count annually, authors with multiple publications, and mutual citations between papers.

Top Journals by Paper Count: This query identifies the top 5 journals for each publication year based on the number of papers published

```
SELECT *
FROM (
  SELECT
    Publication_Year,
    Journal_Name,
    COUNT(*) AS Paper_Count,
    RANK() OVER (PARTITION BY Publication_Year ORDER BY COUNT(*) DESC) AS
Rank_In_Year
  FROM Paper_Basic_Info_Table
  GROUP BY Publication_Year, Journal_Name
) AS RankedJournals
WHERE Rank_In_Year <= 5;
```

Authors with ≥5 Papers: This query identifies authors who have published five or more papers.

```
SELECT
   Author_Name,
   COUNT(*) AS Paper_Count
FROM
   Paper_Author_Relation_Table
GROUP BY
   Author_Name
HAVING
COUNT(*) >= 5;
```

Mutual Citation Count: This query was originally intended to identify paper pairs with mutual citations of at least 3 times, as specified in the task. However, upon analyzing the data, it was found that the maximum number of mutual citations between any paper pairs in the dataset is 2. Therefore, the query was adjusted to identify paper pairs with **at least 2 mutual citations** instead of 3, as no paper pair meets the 3-citation threshold. This adjustment ensures that the query reflects the actual data available, providing a meaningful and accurate result.

```
SELECT
   LEAST(a.Citing_Paper_ID, a.Cited_Paper_ID) AS Paper_A,
   GREATEST(a.Citing_Paper_ID, a.Cited_Paper_ID) AS Paper_B,
   COUNT(*) AS Mutual_Citation_Count
FROM
   Citation_Relation_Table AS a
JOIN
   Citation_Relation_Table AS b
   ON a.Citing_Paper_ID = b.Cited_Paper_ID
  AND a.Cited_Paper_ID = b.Citing_Paper_ID
WHERE
   a.Citing_Paper_ID <> a.Cited_Paper_ID
GROUP BY
   LEAST(a.Citing_Paper_ID, a.Cited_Paper_ID),
   GREATEST(a.Citing_Paper_ID, a.Cited_Paper_ID)
```

```
HAVING
  COUNT(*) >= 2;
```

3.2.2. Result

3.2.2.1. Top 5 journals by paper count annually

IEEE ACCESS consistently ranks as the top journal by paper count across all five years, starting with **116 papers in 2020** and remaining the highest in subsequent years, although with a decreasing trend in total papers published (down to **59 papers in 2024**). This indicates the journal's dominance and consistency in the field over the years.

FINANCE RESEARCH LETTERS and **IEEE INTERNET OF THINGS JOURNAL** are notable for maintaining strong positions in the top 5, consistently ranking high, although their paper counts fluctuate. Specifically, **FINANCE RESEARCH LETTERS** shows growth in 2021 (**28 papers**) but sees a drop in 2024, while **IEEE INTERNET OF THINGS JOURNAL** also sees steady output over the years, peaking at **47 papers in 2022**.

SENSORS remains a prominent journal in this period, securing a place in the top 5 every year with relatively stable paper counts, particularly in **2022** where it ranked **2nd** with **57 papers**.

Key Observations:

- **Declining paper counts**: While **IEEE ACCESS** dominates, the overall paper count has declined over the years (from **116 papers in 2020** to **59 in 2024**), possibly reflecting changing trends in research output or a shift to other platforms.
- **Emerging journals**: **SUSTAINABILITY** and **APPLIED SCIENCES-BASEL** show increasing prominence, especially in 2022 and 2023, indicating a rise in interest in these journals within the blockchain or related fields.
- **Stable high-ranking journals**: **FINANCE RESEARCH LETTERS**, **SECURITY AND COMMUNICATION NETWORKS**, and **IEEE INTERNET OF THINGS JOURNAL** are consistent contenders in the top ranks, contributing to ongoing research in related fields like finance, communication networks, and the Internet of Things.

This analysis shows a pattern of **stable high-output journals** in blockchain and related fields, with some journals experiencing growth and others a slight decline. The dominance of **IEEE ACCESS** and **FINANCE RESEARCH LETTERS** highlights the central role these journals play in research dissemination within their respective domains. The appearance of **newer journals** like **SUSTAINABILITY** and **CLUSTER COMPUTING** indicates evolving research interests in emerging areas.

Table 2. Top 5 Journals by PaperCcount Annually

Publication_Year	Journal_Name	Paper_Count	Rank_In_Year
2020	IEEE ACCESS	116	1
2020	FINANCE RESEARCH LETTERS	24	2
2020	FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE	21	3
2020	SENSORS	21	3
2020	IEEE INTERNET OF THINGS JOURNAL	16	5
2021	IEEE ACCESS	78	1
2021	FINANCE RESEARCH LETTERS	28	2
2021	SECURITY AND COMMUNICATION NETWORKS	28	2
2021	IEEE INTERNET OF THINGS JOURNAL	27	4
2021	SENSORS	25	5
2022	IEEE ACCESS	93	1
2022	SENSORS	57	2
2022	FINANCE RESEARCH LETTERS	50	3
2022	IEEE INTERNET OF THINGS JOURNAL	47	4

2022	SUSTAINABILITY	30	5
2023	IEEE ACCESS	80	1
2023	FINANCE RESEARCH LETTERS	37	2
2023	SENSORS	37	2
2023	IEEE INTERNET OF THINGS JOURNAL	37	2
2023	APPLIED SCIENCES-BASEL	34	5
2023	ELECTRONICS	34	5
2024	IEEE ACCESS	59	1
2024	IEEE INTERNET OF THINGS JOURNAL	21	2
2024	FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE	20	3
2024	FINANCIAL INNOVATION	19	4
2024	CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS	15	5
2024	INTERNATIONAL REVIEW OF FINANCIAL ANALYSIS	15	5

3.2.2.2. Authors with ≥5 papers.

There are 437 authors with more than 5 publications in the list. The table below highlights the most prolific authors. Some notable names include:

- Salah, Khaled: 46 papers
- Tanwar, Sudeep: 45 papers
- Choo, Kim-Kwang Raymond: 39 papers

Table 3. Top Authors with ≥5 papers.

Author_Name	Paper_Count
Salah, Khaled	46
Tanwar, Sudeep	45
Choo, Kim-Kwang Raymond	39
Jayaraman, Raja	38
Zheng, Zibin	36
Zhu, Liehuang	35
Bouri, Elie	34
He, Debiao	29
Kumar, Neeraj	28
Corbet, Shaen	27
Yaqoob, Ibrar	24
Javaid, Nadeem	21

3.2.2.3. Paper pairs with mutual citations ≥2 times

Table 4. Paper pairs with mutual citations ≥3 times

Paper_A	Paper_B	Mutual_Citation_Count
WOS:000605138700003	WOS:000991216000001	2
WOS:000952017800020	WOS:000965433900001	2
WOS:001200424700002	WOS:001218887000001	2

4. Building A Directory Ontology

Ontology design in Protégé facilitates the creation of a structured knowledge model for a specific domain. For Blockchain Literature Metadata, **classes** such as Paper, Author, Institution, Theme, and CitationRelation are defined and interconnected via **Object Properties** (e.g., writtenBy between Paper and Author).

Key Components in Protégé:

- **Classes** define the primary concepts within the domain.
- **Object Properties** establish relationships between these classes (e.g., Paper has a writtenBy relationship with Author).
- **Individuals** are specific instances of these classes, such as Paper1 or Author1.

Example of Relationships in the Ontology:

The ontology below illustrates the classes and their relationships within the Blockchain Literature Metadata, demonstrating how an **Author** is linked to a **Paper** via **Paper-Author**, how a **Paper** is categorized by its **Theme** through **Paper-Theme**, and how a **Paper** is associated with an **Institution** via **Paper-Institution**. Additionally, it shows how a **Paper** is published in a **Journal** through **Paper-Journal** and how citation relationships between papers are captured through **Citation Relation-Paper**.

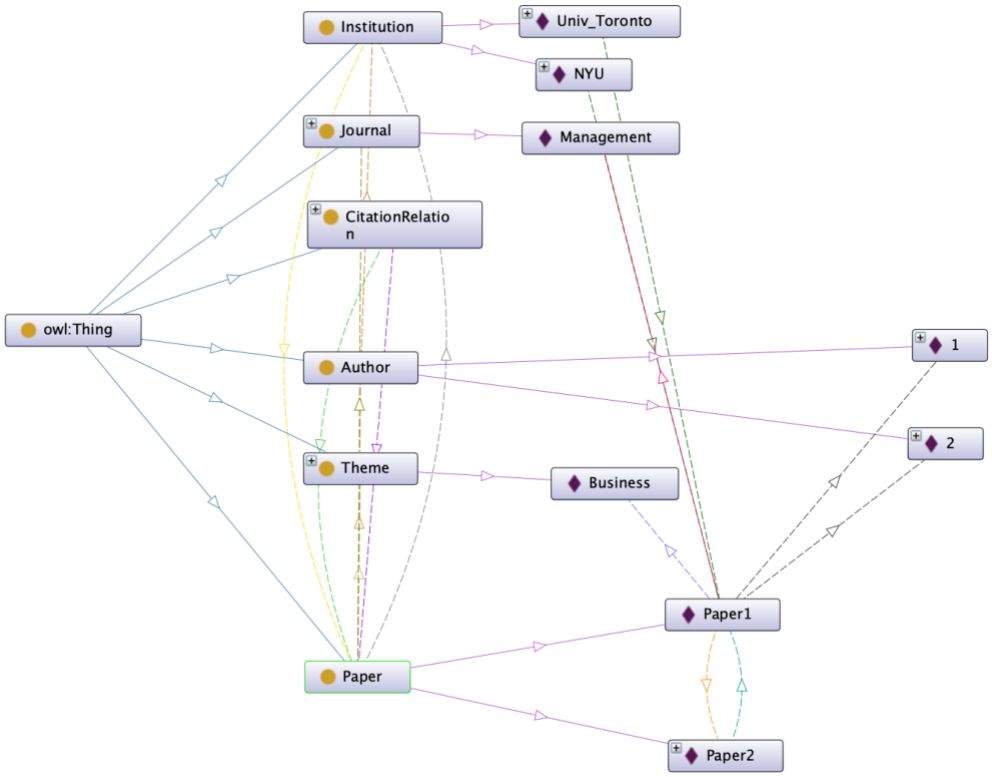

Figure 8. Example Ontology of Blockchain Literature Metadata

5. Analyzing Author Collaboration Networks

5.1. Preparing Network Data

- Nodes file data includes: Author name, address
- Edges file data includes: Source, Tartget, Main Category, Time, Collaboration Count

Source	Target	Main_Theme	Time	Collaboration_Count
Gans, Joshua S.	Halaburda, Hanna	Management	2024	1
Hang, Lei	Kim, BumHwi	Biotechnology & Applied Microbiology	2021	1
Hang, Lei	Kim, KyuHyung	Biotechnology & Applied Microbiology	2021	1
Hang, Lei	Kim, DoHyeun	Biotechnology & Applied Microbiology	2021	1
Dong, Huidong	Liu, Gao	Computer Science, Information Systems	2022	1
Liu, Gao	Yan, Zheng	Computer Science, Information Systems	2022	1
Liu, Gao	Zhou, Xiaokang	Computer Science, Information Systems	2022	1
Liu, Gao	Shimizu, Shohei	Computer Science, Information Systems	2022	1
Badri, Nedia	Nasraoui, Leila	Computer Science, Information Systems	2024	1
Badri, Nedia	Saidane, Leila Azouz	Computer Science, Information Systems	2024	1
Ruj, Sushmita	Sengupta, Jayasree	Computer Science, Information Systems	2023	1
Das Bit, Sipra	Sengupta, Jayasree	Computer Science, Information Systems	2023	1
DeFranco, Joanna	Kassab, Mohamad	Computer Science, Information Systems	2021	1
Kassab, Mohamad	Malas, Tarek	Computer Science, Information Systems	2021	1
Kassab, Mohamad	Laplante, Phillip	Computer Science, Information Systems	2021	1
Destefanis, Giuseppe	Kassab, Mohamad	Computer Science, Information Systems	2021	1
Graciano Neto, Valdemar Vicente	Kassab, Mohamad	Computer Science, Information Systems	2021	1
Azbeg, Kebira	Fetjah, Laila	Medical Informatics	2021	1
Fetjah, Laila	Ouchetto, Ouail	Medical Informatics	2021	1
Andaloussi, Said Jai	Fetjah, Laila	Medical Informatics	2021	1

Figure 9. Example of Edge File imported to Gephi

5.2. Building the Network with Gephi

Designing Network Layout: Layout, Node Size, Color by Theme

Layout: Forte Atlas 2

Edge Color by theme

- Pink: Computer Science, Information Systems (33,13%)
- Green: Business, Finance (7,41%)
- Blue: Computer Science, Hardware & Architecture (6,15%)
- Black: Computer Science, Artificial Intelligence (4,08%)
- Orange: Engineering, Electrical & Electronic (3,53%)
- Dark Pink: Economics (3,43%)
- Mint: Computer Science, Theory & Methods (3,21%)
- Grey: Remaining Category

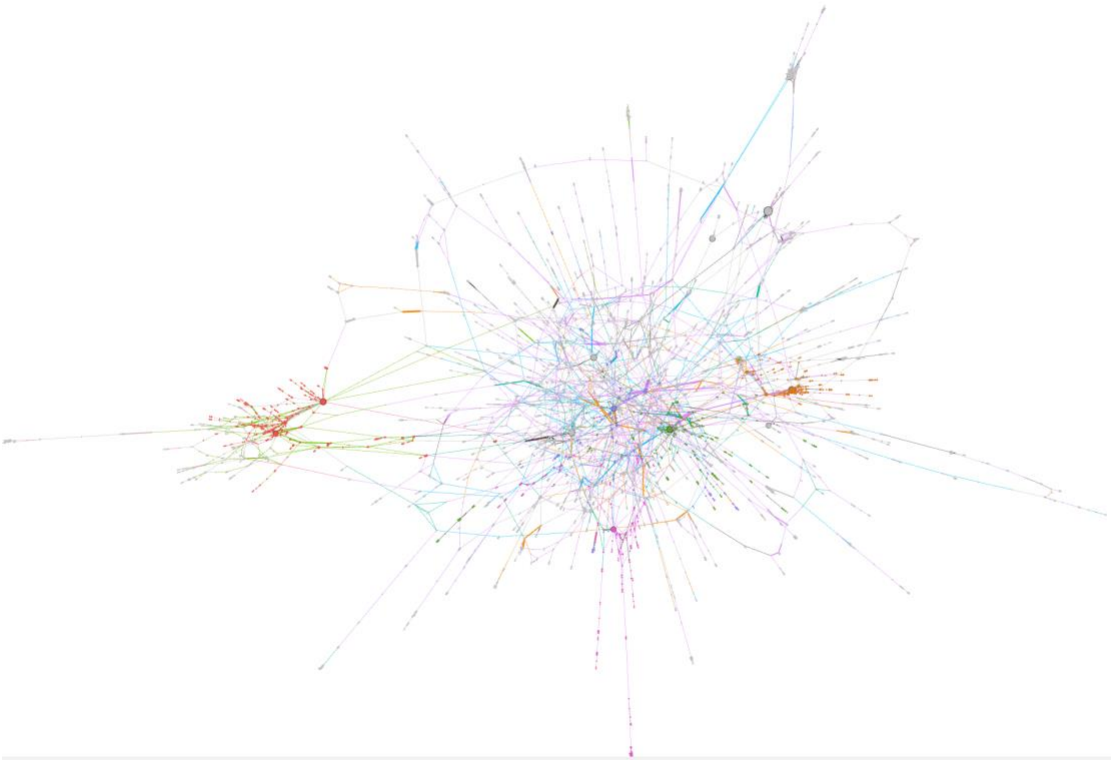

Figure 10. Blockchain Co-authorship Network Graph

(Source: Gephi)

5.3. Analyzing Network Metrics

In this section, we present the results of various network metrics used to analyze the structure of the co-authorship network within the blockchain research field. These metrics provide insights into the importance, connectivity, and structural properties of authors within the network.

5.3.1. Degree Centrality

Degree Centrality measures the number of direct connections a node (author) has in the network. In this study, Degree Centrality was used to identify the most influential authors within

the blockchain research domain, as those with higher degrees are more connected and influential in the collaborative network.

- The author **Khan, Abdullah Ayub** was found to have the highest degree centrality with **47 connections**, highlighting his central role in the network. This suggests that he collaborates frequently with many authors in the blockchain field, indicating his significant influence on the development of research in this area.
- The overall average degree centrality across all authors in the network was **1.945**, indicating a moderately connected network. This suggests that while many authors are involved in the research process, there are key authors who dominate the collaboration space.

Figure 11. Top Authors' Degree Centrality

Table 5. Top Authors' Degree Centrality

Author	Degree
Khan, Abdullah Ayub	47
Tanwar, Sudeep	37
Zheng, Zibin	34
Bouri, Elie	32

5.3.2. Betweenness Centrality

Betweenness Centrality is a measure of a node's role as a bridge within the network. Nodes with high betweenness centrality have the potential to control the flow of information between other nodes in the network, acting as intermediaries or brokers.

- Among the authors analyzed, **Bouri, Elie** emerged with the highest Betweenness Centrality score of **3,136,041.07**. This indicates that **Bouri, Elie** serves as a central bridge in the network, connecting distinct research clusters. Such authors are crucial in facilitating information flow and collaboration across different subfields, highlighting their influence in the blockchain research community.
- The presence of authors with high betweenness centrality is critical for understanding how ideas and research are transferred across different communities and regions in the blockchain field.

Table 6. Top Authors' Betweenness Centrality

Author	Betweenness Centrality
Bouri, Elie	3136041.065599052
Choo, Kim-Kwang Raymond	3018225.483694995
Zheng, Zibin	2865764.951227124
Wang, Lei	1621493.8697239496
Guizani, Mohsen	1563902.9378207542

5.4.3. Clustering Coefficient

The **Clustering Coefficient** measures the extent to which nodes in a network tend to cluster together. High clustering indicates that authors who collaborate with a common set of co-authors are more likely to be closely interconnected, forming tight-knit communities within the broader network.

- The **Clustering Coefficient** in the blockchain research network was found to be **0.127**, suggesting a relatively low level of clustering. This indicates that while there are some

tightly-knit groups of authors collaborating, the blockchain research network is still somewhat dispersed with authors collaborating across different subgroups or topics.

- A lower clustering coefficient means that the network is more loosely connected, with fewer closed triangles between authors. However, this also suggests a potentially more diverse set of collaborations, where ideas and research can cross boundaries more easily.

5.3.4. Network Diameter

The **Network Diameter** is the longest shortest path between any two nodes in the network. It provides an indication of the "diameter" or maximum distance between the farthest nodes in the network, reflecting the overall connectivity and efficiency of the network.

- The **Network Diameter** of the blockchain research network was calculated to be **32**, meaning the longest shortest path between two authors in the network consists of 32 steps. This suggests that, while the network is relatively large, it is still fairly compact compared to other large academic networks.
- The **Network Radius**, which is the shortest path from the most central node to any other node, was **1**, indicating that the network is well-connected with a central core.
- The **Average Path Length** in the network was found to be **10.45**, indicating that, on average, it takes about 10.45 steps for one author to reach another in the network. This suggests a moderate level of efficiency in information spread, with most authors being relatively close to each other in terms of collaborative distance.

5.4. Community Detection

Community sizes: Approximately 2,352 communities were detected using the modularity algorithm. The chart below illustrates the five communities with the highest number of authors.

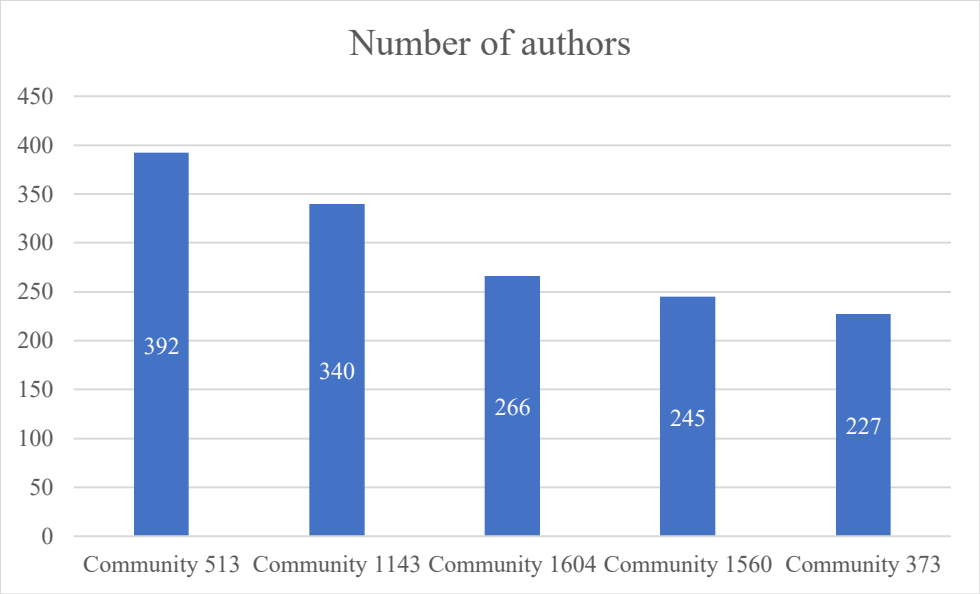

Figure 12. Top 5 Community Size

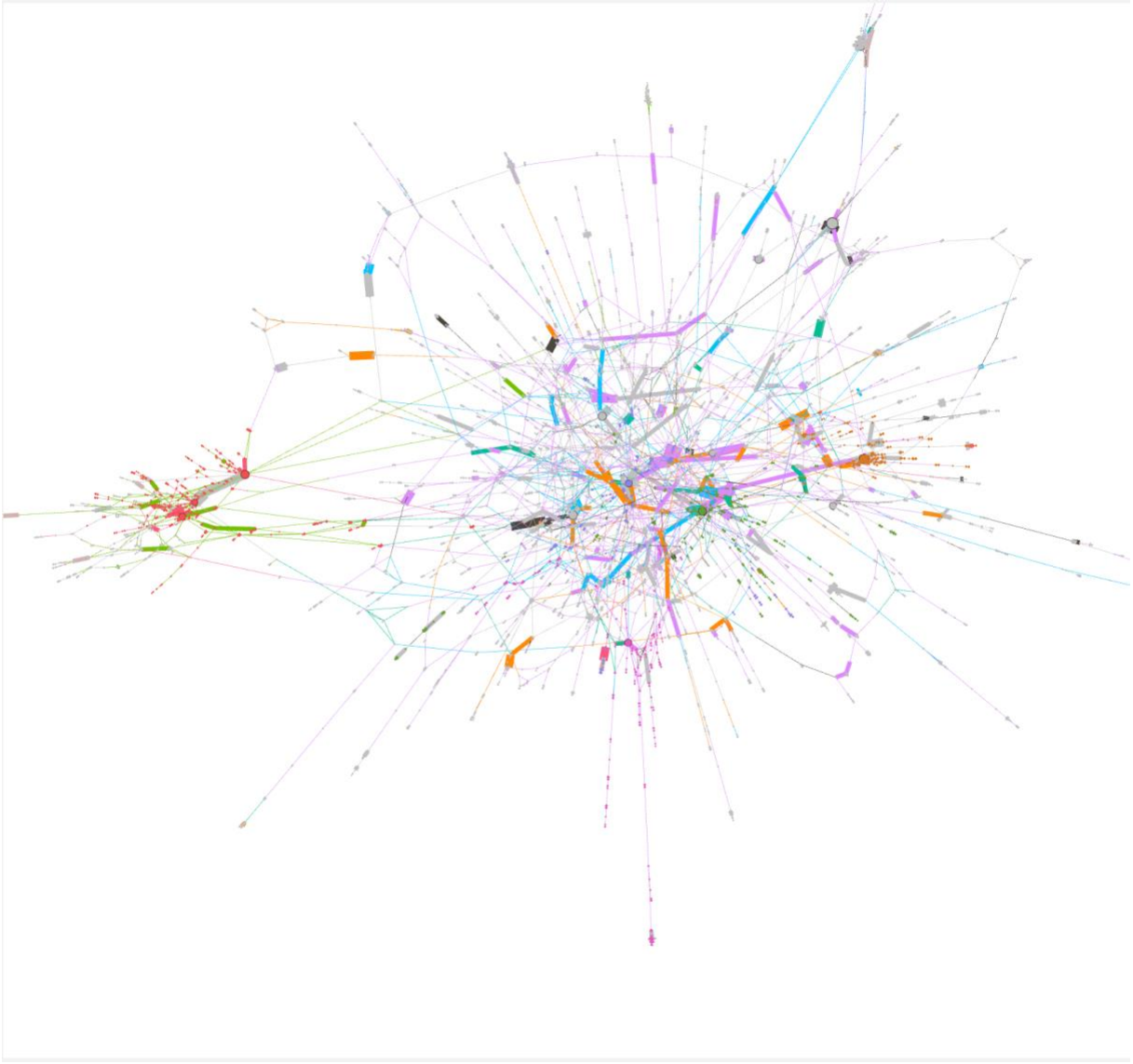

Figure 13. Topic-Based Co-authorship Network in Blockchain Research (Edge Color = Topic, Node Color = Community, Edge Thickness = Collaboration Strength)

Note:

Line Color

- Pink: Computer Science, Information Systems (33,13%)
- Green: Business, Finance (7,41%)
- Blue: Computer Science, Hardware & Architecture (6,15%)
- Black: Computer Science, Artificial Intelligence (4,08%)
- Orange: Engineering, Electrical & Electronic (3,53%)

- Dark Pink: Economics (3,43%)
- Mint: Computer Science, Theory & Methods (3,21%)
- Grey: Remaining Category

Community and Topic Linkage Analysis

- **Strong Inter-Community Connections**: The thick and vivid-colored edges represent strong collaborative ties between major communities. These large clusters often share similar research topics. For instance, communities like **Community 513 (red nodes)** are associated with **Business and Finance**, while communities such as **1143 (green nodes)**, **1604 (purple nodes)**, and **1560 (pink nodes)** are centered around **Computer Science** themes.
- **Reassessment of Linkage Themes**: Upon closer inspection, it appears that some purple-colored nodes (e.g., Community 1604) are frequently connected via **thick orange lines**, which correspond to **Engineering, Electrical & Electronic** topics—suggesting that this community may not be primarily focused on Computer Science as initially assumed, but rather on engineering-oriented blockchain applications.
- **Smaller Communities**: Smaller and less prominent communities (indicated by lighter-colored and smaller nodes) exhibit fewer interconnections. The thin and faint-colored edges connecting them suggest minimal overlap in research themes or collaboration. These may represent niche research areas or isolated institutional efforts.

Top 5 Community Analysis

Figure 14. Community 513

Community 513 (red nodes) exhibits strong internal connectivity. The nodes are densely clustered and connected by thick edges, indicating close collaboration among members. Larger nodes represent key authors with influential roles in the community.

Connections to other communities (grey nodes) are present but less prominent, suggesting that Community 513 primarily focuses on internal collaboration. However, a few weaker links to other clusters may imply some degree of interdisciplinary interaction.

Connectivity insights:

- **Tightly knit community**: High density and strong internal collaboration.
- **Key authors**: Large nodes indicate central, influential researchers.
- **Limited external links**: Mostly internal focus with minor cross-community ties.

Thematic focus:

- **Strong emphasis on Business and Finance**: Suggested by dominant edge colors.
- **Intra-field cooperation**: Authors in this domain collaborate closely.
- **Weak cross-topic interaction**: Some links to other fields (e.g., engineering) exist but are less significant.

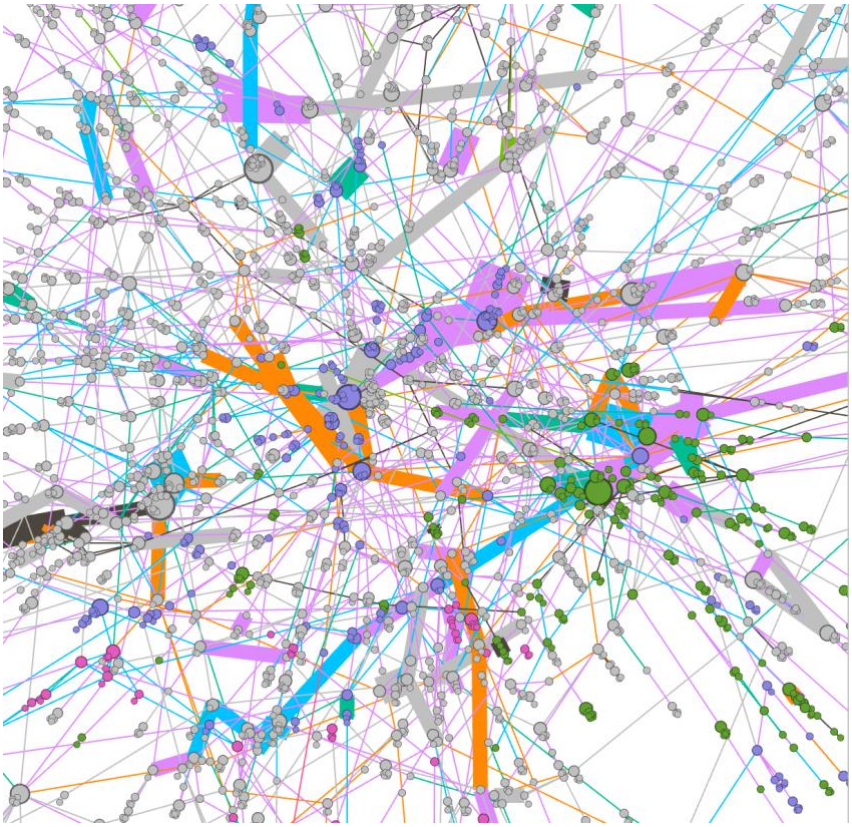

Figure 15. Community 1604

Community 1604 (purple nodes) shows strong internal connectivity but exhibits a highly **diverse and interdisciplinary thematic distribution**. While the core concentration remains in **Computer Science**—especially in Hardware & Architecture, Theory & Methods, and Information Systems—there is significant overlap with fields such as **Engineer Electrical (Orange Line)**, **Mathematics**, **Telecommunications**, **Business**, and even **Green & Sustainable Science**.

This pattern suggests that Community 1604 acts as a **bridge across multiple research domains**, reflecting a collaborative structure that supports cross-disciplinary innovation in blockchain-related studies.

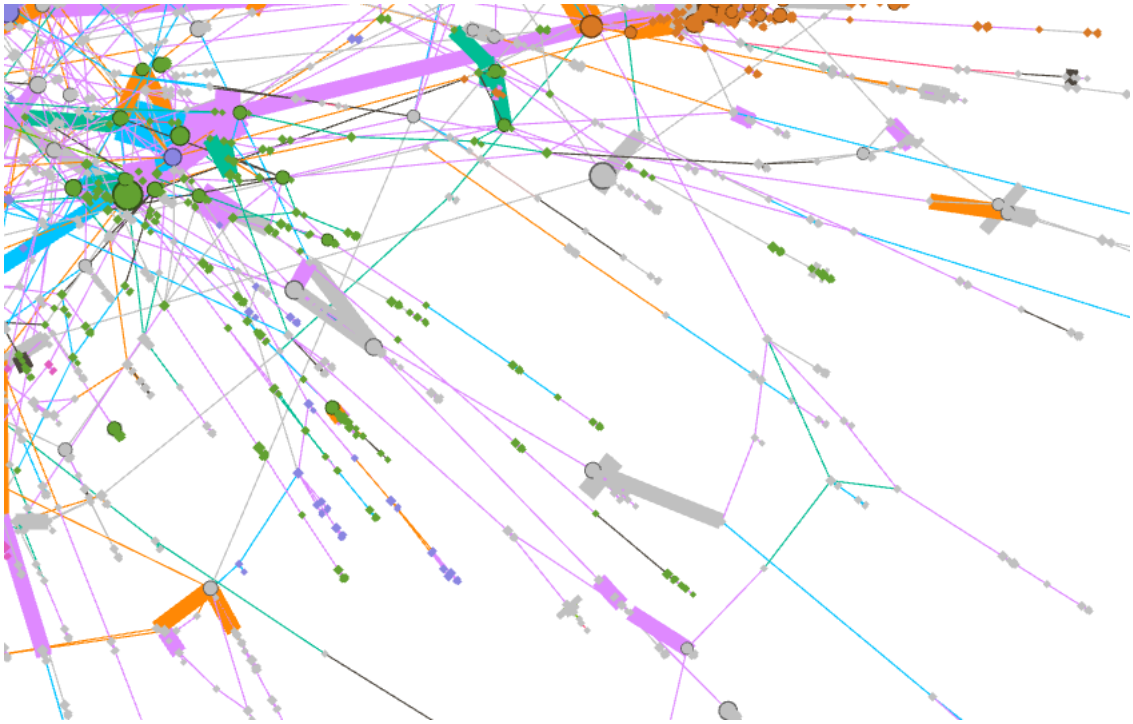

Figure 16. Community 1143

Community 1143 (green) is a highly interconnected and interdisciplinary author group, bridging core areas of computer science with business, engineering, and emerging application domains.

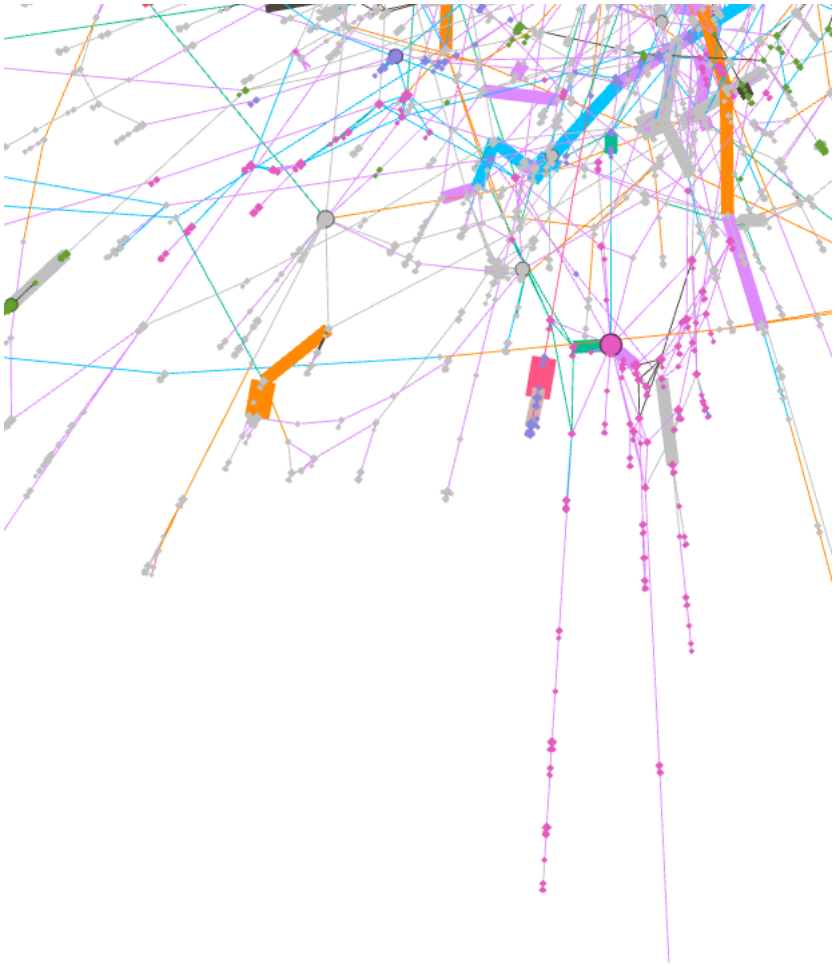

Figure 17. Community 1560

Community 1560 (pink) is a technically driven cluster, with strong internal collaboration centered on automation, computer systems, and sustainable engineering applications.

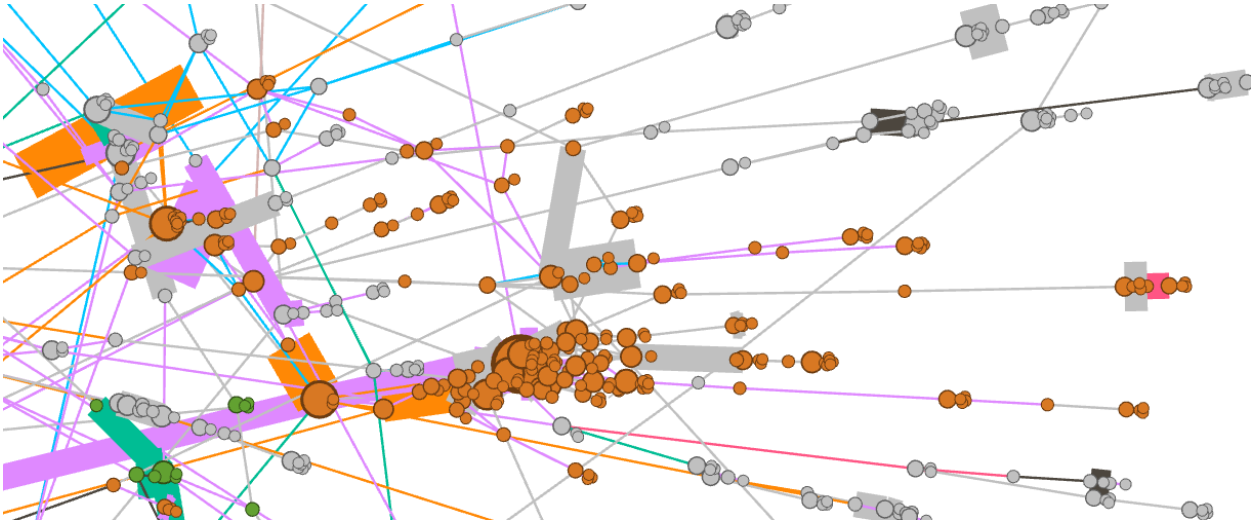

Figure 18. Community 373

Community 373 (orange nodes) shows **high node density and compact distribution**, with many nodes closely clustered and interconnected. A few larger central nodes indicate influential authors, while thick orange edges reflect tight internal collaboration.

Thematically, this group is centered around **Computer Science, Information Systems (pink line)**, strongly linked with **Construction & Building Technology**, **Civil and Electrical Engineering (orange line)**, and **Green & Sustainable Science & Technology**. The presence of **Energy & Fuels**, **Mathematics**, and **Economics** suggests a **highly interdisciplinary community**, focused on the **application of computational systems in construction, energy, and environmental domains**.

5.5. Temporal Analysis

Table 7. Network Index Summary Table 2020-2024

Year	Modularity	Number of Community	Average Degree	Density	Diameter	Average Path Length	Weakly Connected Components
2024	0.994	638	1.636	0.001	15	4.35	637
2023	0.991	1034	1.684	0.000	32	12.51	1026

2022	0.992	952	1.714	0.000	25	11.27	942
2021	0.991	728	1.688	0.000	26	6.94	723
2020	0.990	570	1.654	0.001	22	7.93	564

Table 8. Table of Highest Degree Centrality and Betweenness Centrality Indexes by Year

Year	**Degree Centrality**	**Name**	**Betweenness Centrality**	**Name**
2024	12	Zhang, Lejun	2670.0	Guizani, Mohsen
2023	15	Leng, Jiewu	141131.41	Liu, Zhenguang
2022	20	Khan, Abdullah Ayub	105039.95	Wang, Yuntao
2021	27	Asuncion, Francis	7397.0	Li, Yi
2020	17	Corbet, Shaen	14427.5	Han, Gang

Overview of Network Evolution (2020–2024)

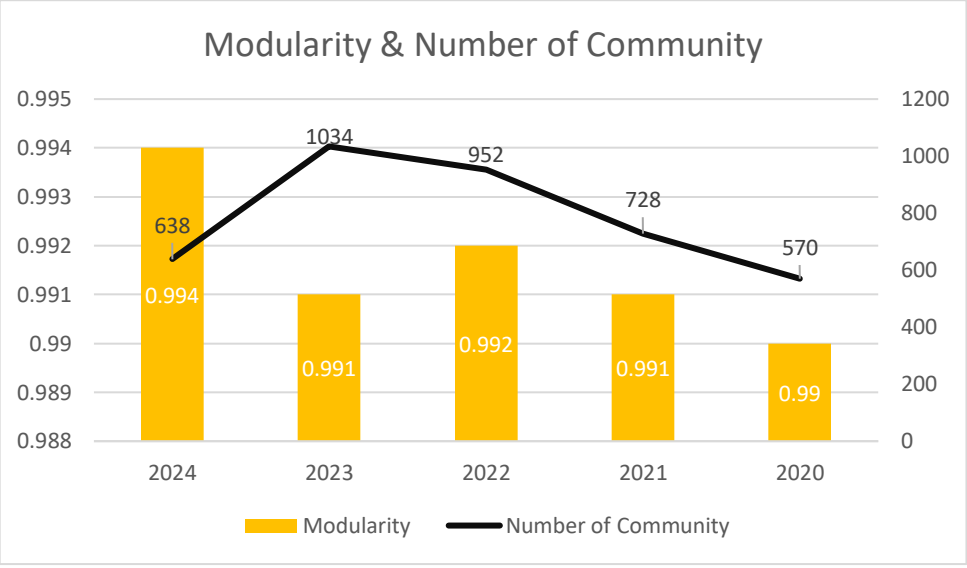

Figure 19. Modularity and Number of Community from 2020 - 2024

Key Trends and Interpretation

a. 2020–2021: Emergence and Initial Diversification

- **Modularity** rose slightly (0.990 → 0.991) and **communities increased** (570 → 728), reflecting early diversification of blockchain research into subtopics (e.g., consensus protocols, tokenization, privacy).
- **Low average degree** and **high number of weakly connected components** (>560) indicate a still-fragmented network with minimal inter-group collaboration.
- **Implication**: This aligns with blockchain's academic phase of theoretical exploration and foundational work (e.g., security models, PoW/PoS analysis).

b. 2022–2023: Explosive Growth and Fragmentation

- The number of **communities peaked in 2023** (1034), while **modularity remained high (~0.991–0.992)**.
- **Network diameter** and **average path length** surged (Diameter: 25→32; Path Length: 11.27→12.51), and weakly connected components exceeded 1000.
- **Interpretation**: Blockchain research became more fragmented, possibly due to the emergence of specialized domains like **DeFi**, **NFTs**, **cross-chain protocols**, and **Web3 governance**.
- **Academic dynamic**: Rapid expansion of niche topics created silos, with researchers working within tightly scoped domains and limited cross-community interaction.

c. 2024: Community Consolidation and Integration

- **Modularity increased to 0.994** (highest), yet **number of communities dropped sharply** (1034 → 638), and **diameter reduced significantly** (32 → 15).
- **Average path length dropped** (12.51 → 4.35), suggesting a denser and more integrated network.
- **Fewer disconnected groups** (1026 → 637) and slight increase in **density**.
- **Implication**: This signals a turning point—previously isolated communities began to **merge or collaborate**, likely due to:

- Integration of **blockchain with real-world systems** (supply chains, healthcare, finance).
- Maturity of DeFi/NFTs shifting research from hype to interdisciplinary optimization.
- Rise of infrastructure standardization (e.g., interoperability, layer-2 scaling) requiring collaboration across domains.

Linking to Blockchain Research Dynamics

Research Phase	Network Trend Observed	Blockchain Field Dynamics
2020–2021 (Formative)	Fragmented communities, small path lengths, moderate modularity	Foundational theories, consensus models, cryptographic primitives
2022–2023 (Boom & Divergence)	Peak number of communities, rising path lengths, large disconnected components	Explosion of subfields (DeFi, NFTs, DAOs); thematic isolation
2024 (Convergence)	Fewer communities, higher modularity, denser network, shorter paths	Interdisciplinary applications (blockchain + AI/IoT), collaborative maturity

Conclusion

From 2020 to 2024, the blockchain academic network transitioned from an exploratory phase with siloed communities to a more **integrated and collaborative ecosystem**. This mirrors real-world blockchain dynamics:

- Early focus on core protocols and cryptographic mechanisms.
- Mid-phase expansion into niche, hype-driven subfields.
- Recent convergence around cross-domain applications and infrastructure standardization.

These structural shifts highlight blockchain's **evolution from innovation to integration**—a critical insight for understanding both its scholarly impact and future trajectory.

6. Conclusion

This study presents a comprehensive framework for organizing, modeling, and analyzing academic metadata to uncover collaboration patterns and thematic structures in blockchain research from 2020 to 2024. By integrating Excel-based preprocessing, relational database modeling, ontology development, and network analysis with Gephi, the research reveals a dynamic evolution of the scholarly landscape—from early-stage fragmentation to increasing interdisciplinary convergence. The time-series analysis confirms that while initial years featured siloed collaborations and topic-specific clusters, recent trends reflect a shift toward integrated research efforts, driven by real-world blockchain applications and technological maturity. These findings contribute valuable insights into the structural development and future direction of blockchain research networks.