Article

Not peer-reviewed version

An analysis of decentralised systems in environment-related projects: theoretical and practical perspective

Matteo Vaccargiu * and Roberto Tonelli

Posted Date: 16 July 2024

doi: 10.20944/preprints202407.1241.v1

Keywords: Blockchain; Earth; Energy; Sustainability; Bibliometric Analysis; Topic Analysis

Article

An Analysis of Decentralised Systems in Environment-Related Projects: Theoretical and Practical Perspective

Matteo Vaccargiu [1,2,†,‡], **Roberto Tonelli** [1,‡]

[1] University of Cagliari; {matteo.vaccargiu, roberto.tonelli}@unica.it
[2] University of Camerino; matteo.vaccargiu@unicam.it
* Correspondence: matteo.vaccargiu@unica.it
† Current address: Palazzo delle Scienze, via Ospedale 72, 09121 Cagliari, Italy
‡ These authors contributed equally to this work.

Abstract: The growing interest in environmental sustainability issues and, at the same time, the advantages offered by blockchain technology have strong connections between these two topics. This study explores the application of blockchain technology across various environmental domains, such as air quality, climate change impacts, and resource management. The research utilised a dual approach, combining a bibliometric analysis with VOSviewer and a topic analysis using BERT models to assess the discourse within both scientific literature extracted from Scopus and practical blockchain projects obtained from GitHub. The findings reveal that food security, energy, and sustainable agriculture are predominant topics in academic discussions, with a noticeable increase in focus from 2017 onwards. Practical projects are focused on transparent tracking and decentralised management. The overlap between academic and practical spheres is evident in the shared focus on energy and environmental management, demonstrating blockchain's growing role in addressing global environmental challenges. The study underscores the importance of integrating theoretical research with practical implementations to harness blockchain's full potential in promoting sustainable environmental practices.

Keywords: blockchain; earth; energy; sustainability; bibliometric analysis; topic analysis

---

1. Introduction

In recent years, blockchain technology and Distributed Ledger Technology (DLT) have emerged as significant tools capable of revolutionising various sectors by providing enhanced transparency, security, and efficiency [1]. These technologies are particularly relevant in environmental contexts, where they can significantly contribute to monitoring, reporting, and verification of data in applications ranging from air quality management to biodiversity conservation. Their application directly supports the achievement of the Sustainable Development Goals (SDGs), particularly those related to environmental protection and sustainable resource management [2]. The urgency of addressing environmental challenges such as air pollution, climate change, water scarcity, and the sustainable management of natural resources is more pronounced than ever before.

Blockchain technology, defined as a sharing register that ensures the immutability and transparency of data transactions [3], offers a robust framework for tackling complex environmental issues. Its decentralised nature prevents tampering and ensures the integrity of environmental data [4]. This is crucial for tracking atmospheric changes, managing the environmental impacts of climate dynamics, and ensuring the sustainable use of water resources. Furthermore, blockchain can facilitate enhanced resource management strategies, aid in disaster risk reduction, and support sustainable urban development by enabling more effective coordination and management of environmental policies and practices [5].

This paper aims to explore the application of blockchain technology across several environmental domains, including air quality and pollution control, climate change impacts, water quality management, and the interplay between land use and environmental health. Based on a bibliometric and topic modelling approach, we tried to answer the following research questions:

RQ1: What are the main discussion topics within **scientific literature** regarding the use of blockchain in critical areas such as earth sciences, climate change, and environmental health? How have these discussions evolved over time?

RQ2: What themes emerge from **practical blockchain projects** in these fields, and how do they develop over time?

RQ3: Is there an **overlap** between the themes explored in **academic research** and those implemented in **practical projects**?

To answer these questions, we performed a bibliometric analysis using VOSviewer and a topic analysis with a BERT (Bidirectional Encoder Representations from Transformers) model with the goal of mapping the current landscape of blockchain applications in environmental research and to highlight how this technology is being used to address pressing global challenges. By bridging the gap between theoretical research and practical applications, this paper seeks to inform and inspire stakeholders from various sectors, particularly those without an IT background, about the potential of blockchain technology in promoting environmentally sustainable behaviors. Finally, in order to support global sustainability initiatives, we would like to promote the wider usage and applications of blockchain technology to address environmental problems.

The paper is organised as follows: in Section 2 we present the existing literature connected to this study; then, in Section 3 we discuss in detail the methodology followed to answer the research questions. The results are presented in Section 4, and discussed and validated in Section 5 and 6 respectively. Finally, in Section 7, conclusions and future research developments in this area are discussed.

2. Related Works

The exploration of blockchain technology in environmental and sustainability domains has attracted the attention of the attention of the scientific community, focusing on its application across various critical areas such as climate change, energy sustainability, and environmental management. This section outlines the key contributions from recent literature, comparing them with some of the of the objectives and methodologies of this study.

Jin et al. [6] addressed the integration of blockchain in environmental management frameworks, demonstrating its potential through bibliometric analysis while noting the scarcity of practical implementations. Their work underlines the foundational stages of blockchain applications in this field, which aligns with the preliminary findings of this study through bibliometric mapping. They use VOSviewer for the analysis, as in our work.

O'Donovan et al. [7] conducted an extensive review of blockchain applications within the energy sector, emphasising the gap between theoretical research and practical applications. Their insights into real-world blockchain initiatives offer a critical perspective that complements the practical component of this study, where real-world blockchain projects from GitHub were analysed.

Joshi et al. [8] systematically reviewed the literature on blockchain's impact on sustainable development, linking it to the United Nations Sustainability Development Goals. This study extends their thematic analysis by using VOSviewer and topic modelling to examine how these themes are discussed in recent scientific literature and practical projects.

Popkova et al. [9] explored the conceptual and empirical applications of blockchain for climate change and clean energy, which are in accordance with the areas of interest of this research. The discussion of the role of blockchain in promoting green initiatives and sustainable investments provides a comparative basis for evaluating the results of this study's bibliometric analysis and review of GitHub projects.

Böckel et al. [10] examined the potential of blockchain to support circular economy approaches. Their analysis of the challenges and opportunities mirrors the dual analytical approach of this study, where both academic and practical perspectives are considered to assess blockchain's impact on environmental sustainability.

Furthermore, Gawusu et al. [11] and Wang et al. [12] provided insights into the integration of blockchain with renewable energy sources, noting significant research interest and practical developments in this area. These findings are critical as they align with this study's focus on energy-related topics within the blockchain discourse.

Lastly, Dorfleitner et al. [13] and Arshad et al. [14] contributed empirical and theoretical insights into blockchain applications that specifically target climate protection and sustainability goals. Their discussions on the operational success factors and the strategic implications for policymakers provide a valuable framework for the discussions in this paper.

In summary, while the reviewed literature lays a robust foundation for understanding blockchain's role in sustainability and environmental management, this study contributes an integrated analysis of scientific articles and practical projects, providing a full spectrum of the potential of blockchain in these critical areas. This paper seeks to bridge the identified gaps between theoretical advancements and practical implementations, providing a comprehensive overview of the current state and future potential of blockchain technologies in environmental sustainability.

3. Methodology

This work aims to provide an overview and statistics about topics related to theoretical research and practical applications of blockchain technology in environmental projects. In this section, we offer a detailed look at the dataset used and explain the method we employed to extract and analyse the topics.

3.1. Datasets Overview and Statistics

For the bibliometric mapping analysis and for the topic modeling of the scientific literature, we considered a dataset extracted from Scopus using the following research query: *("blockchain" OR "DLT") AND ("earth" OR "Air quality" OR "pollution" OR "Environmental impacts" OR "climate change" OR "Water quality" OR "Sustainable urban development" OR "Soil system" OR " Natural disasters" OR "human-made disasters").*

Based on that research query, we obtained a set of 1262 documents from 1995 until June 2024, of which 1238 were in English, 17 in Chinese, 5 in Japanese, 2 in Spanish, and 1 in Bosnian.

We considered only the English documents for the analysis. Then we also exclude 112 proceedings data keeping for the analysis only research papers. After we exclude the other six articles without an abstract. We obtain, at the end of the process, 1120 papers.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2024 doi:10.20944/preprints202407.1241.v1

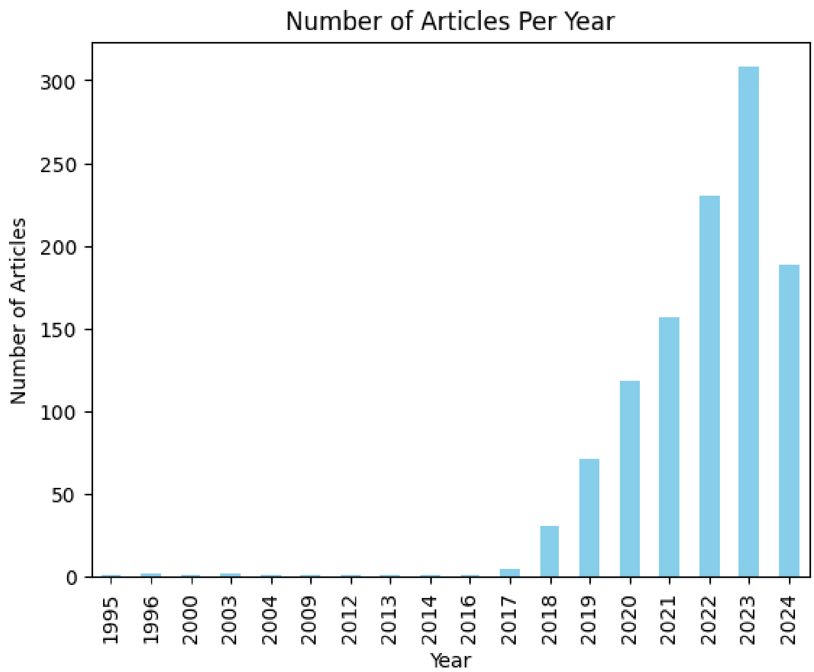


Figure 1. Number of articles published by year.

In Figure 1, we have the distribution of the articles over time. As we can observe from the plot, the trend of interest in these themes has been growing steadily and continuously since 2017, with a peak in 2023. The mean number of citations for these articles is 11.98, and the most cited article is titled "Internet of things (IoT) and the energy sector", cited 460 times.

For the practical project analysis, we considered a dataset extracted from GitHub, which is a web-based platform that provides hosting for software development and version control. For these reasons, it is a good space to research blockchain projects related to topics of sustainability, earth, pollution, water, air, etc. and how developers work on these issues. In particular, we extracted all repositories and all issues that are obtained with the string *"earth+blockchain"*. We thus obtained a dataset of 1000 issues and one of 59 repositories, one of which was eliminated because it was empty and lacked description.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2024 doi:10.20944/preprints202407.1241.v1

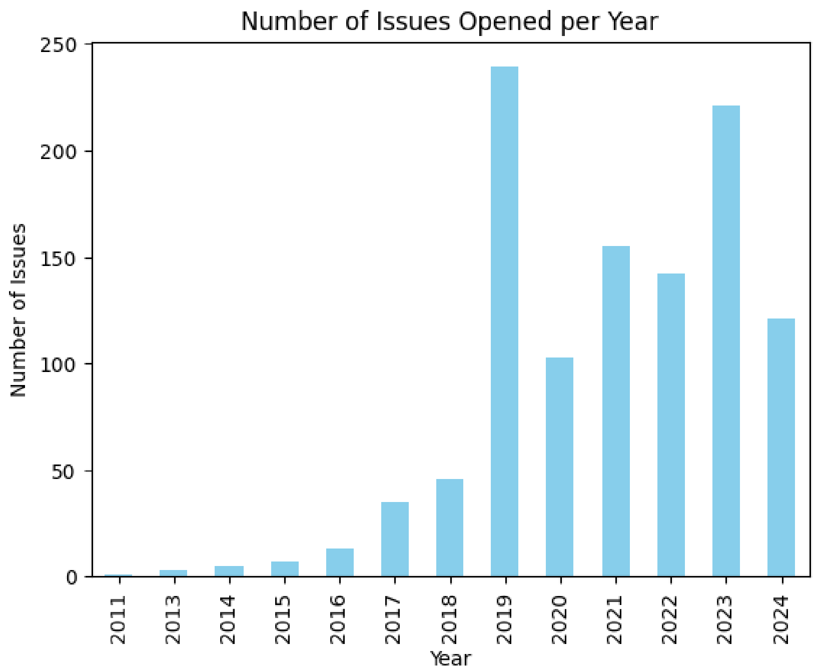


Figure 2. Number of issues opened by year.

In Figure 2, we have the distribution of the opened issues over time. As we can observe from the plot, we have two peaks, one in 2019 and one in 2023. The mean number of citations for these articles is 11.98, and the most cited article is titled "Internet of things (IoT) and the energy sector", cited 460 times. Of these issues, 442 are still open, while 558 have been closed. The mean lifespan of issues is 135.31 days. The mean number of comments on an issue is 24.32, and the most commented one has 1907 comments.

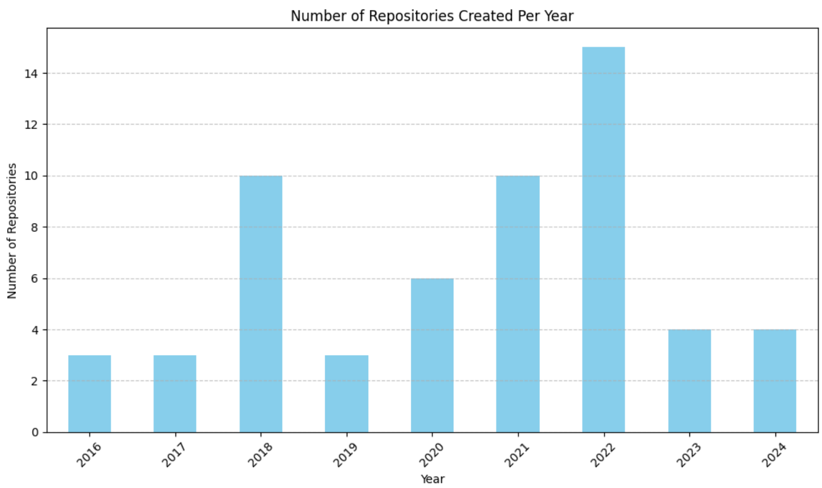


Figure 3. Number of repositories created by year.

In Figure 3, we have the distribution of the created repositories over time. The mean size of these repositories is 8894.78 KB, and the biggest is 156808 KB. Moreover, we have 9 different programming languages used for blockchain projects in the earth field, and the top 3 most used are JavaScript 14 times, TypeScript 4 times, and Solidity 4 times. For this result, it is important to highlight that only 31 repositories declare the programming language used. Finally, the mean number of open issues per repository is 1.57, and the one that has 70 issues open is a content delivery network (CDN) that uses Ethereum and IPFS.

 doi:10.20944/preprints202407.1241.v1

3.2. Theoretical Projects Analysis

The primary descriptive analysis was a bibliometric map on the Scopus data performed using VOSviewer, a software tool for constructing and visualising bibliometric networks [15]. Considering both the titles and abstracts of the paper, we perform both full and binary counting. The full counting helps capture all occurrences without reducing the weight of multiple contributions within the same item. It gives more weight to terms that appear frequently, which is helpful in the study of the influence or prevalence of certain topics. On the other hand, binary counting is useful to avoid overrepresentation of a specific word. This is useful for analysing the breadth of topics in the dataset, and you want to minimize the influence of prolific terms. The results obtained are shown in Figure 4 and Figure 5.

Before analysing the research topics in more detail by applying a natural language processing (NLP) model, we preprocessed the data using Berteley's preprocess function, which is designed to systematically prepare textual data. It accepts a list of documents and performs a series of predefined cleaning and normalization steps on the text. By incorporating these steps, the function ensures that the input text is standardised, potentially enhancing the performance of subsequent NLP tasks.

Then, once the data has been cleaned up, we performed a topic analysis using the BERT (Bidirectional Encoder Representations from Transformers) model [16], which is a pre-trained transformer-based neural network model designed to understand the context of a given text by bidirectionally processing it. Specifically, we used *BERTopic* [1], a model that leverages contextual embeddings from BERT to identify and cluster topics within a collection of text documents. To encode the input text into a fixed-size hidden representation, BERT's architecture incorporates a multi-headed self-attention mechanism and a feed-forward neural network. To predict missing words in a given sentence, the BERT pre-training strategy involves training the model on a sizable corpus of unannotated text. Specifically, in our case, we applied an unsupervised approach, considering two different embedding models. The first one, called *SciBERT*, is trained on a dataset of scientific articles [17], while the second one, called *ClimateBERT*, is trained on a dataset related to climate, sustainability, and environment [18]. This choice is due to the fact that, by doing so, we get topics extracted from two different points of view. In addition, the choice of models was made based on the c_v score measure [19]. After the first process of topic extraction, we employed the feature *"reduce_outliers"* of BERTopic that helped us reduce the unclassified documents, distributing them in clusters based on the class-based TF-IDF (c-TF-IDF).

Table 1. Comparison of Embedding Models

Embedding Model	Num Topics	c_v Score with Outlier	c_v Score Outlier Reduced
SciBERT	20	0.6027	0.5802
ClimateBERT	22	0.5880	0.5811

In Table 1, we have the number of topics and the c_v scores obtained with the two embedding models before and after the outlier reduction.

Once we obtained the list of the topics with the 10 top words associated with each one, we interpreted them for labeling the topics using Chat-GPT 4, followed by a manual check by the authors for validating the results. The effectiveness of the use of Chat-GPT for topic labeling is demonstrated by Colavito et al.[20].

The results obtained are shown in Table 2 and in Figure 6, Figure 7, and Figure 8.

3.3. Practical Projects Analysis

Based on the issues and repositories data extracted from GitHub, we performed a topic analysis based on the title and body of each issue, and we analysed the description and topic for each repository.

[1] https://maartengr.github.io/BERTopic/index.html

 doi:10.20944/preprints202407.1241.v1

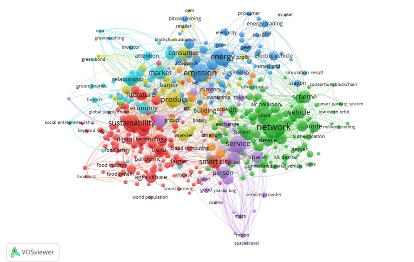
Based on the idea of Vaccargiu et al. [21], for the analysis of the issue we applied *BERTopic* using the bge embedding model, especially in this case we considered the one called *BAAI/bge-reranker-base* because it is trained both in English and Chinese and some issues in our dataset are written in Chinese. At the end of the process, we obtain 20 topics and a c_v score measure of 0.6174. As done previously for scientific articles, wanting to classify all issues, we reduced the outliers by distributing them in the other classes, thus obtaining 19 topics and a c_v score measure of 0.6176. The results obtained are shown in Table 3.

Then we moved on to the repository data, and we performed another topic analysis. Specifically in this case, not getting satisfactory results from the direct application of BERTopic to the description of repositories, probably due to the too short length of many descriptions, we extracted keywords from them using *KeyBERT*. Which is a BERT-based package for keyword extraction. By utilising these models, KeyBERT can efficiently identify and extract the most relevant and contextually significant keywords and keyphrases from a body of text. Especially in our case, we extracted keywords from the repository descriptions and, combined with those provided by some developers, interpreted them using Chat-GPT 4 and human check to understand the topic of each repository. Once we obtained the topics, using a similar approach, we clustered them based on the topics covered. The results obtained are shown in Table 4.

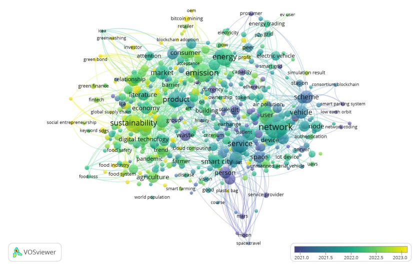
4. Results

This section presents the results of the bibliometric analysis and the interpretation of the resulting topics obtained in the different NLP processes.

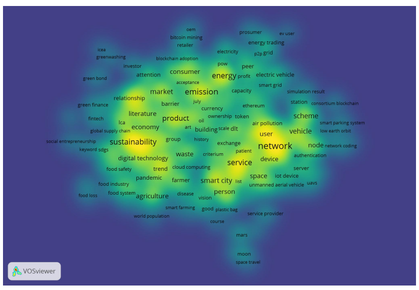
Performing the analysis with VOSviewer, we set the minimum number of occurrences of a term as 10 and considered only the 60% most relevant term. Using the full counting, we selected 532 words, and instead of using the binary counting, we considered 512 items. With the first method, we obtain 7 clusters, 40563 links, and a total link strength of 183361. and the results are shown in Figure 4. Instead of the second one, we have 3 clusters, 37357 links, and a total link strength of 83631. The results obtained are plotted in Figure 5.



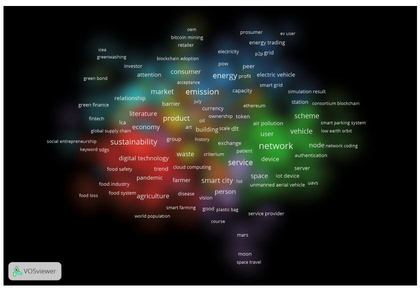
(a) Network Visualization with the full counting method analysis



(b) Overlay Visualization with the full counting method analysis

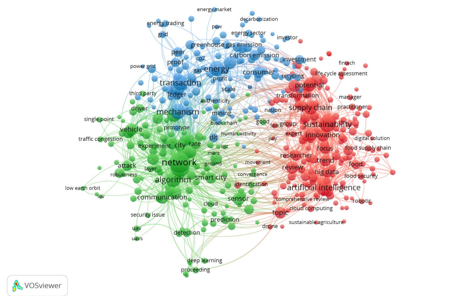


(c) Density Items Visualization with the full counting method analysis

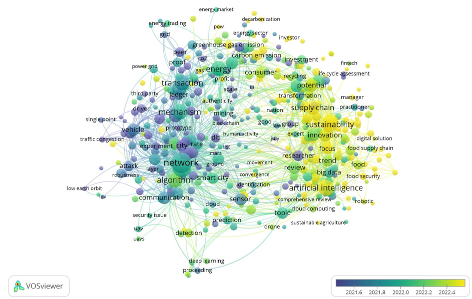


(d) Density Clusters Visualization with the full counting method analysis

Figure 4. VOSviwer map analysis plots using a full counting method.

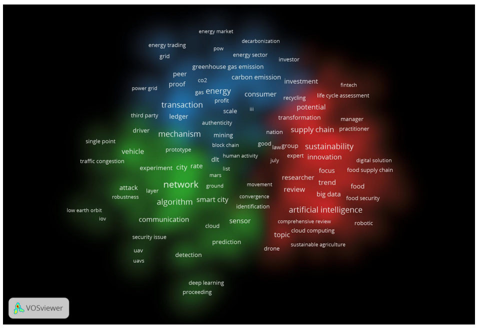


(a) Network Visualization with the binary counting method analysis



(b) Overlay Visualization with the binary counting method analysis

(c) Density Items Visualization with the binary counting method analysis



(d) Density Clusters Visualization with the binary counting method analysis

Figure 5. VOSviwer map analysis plots using a binary counting method.

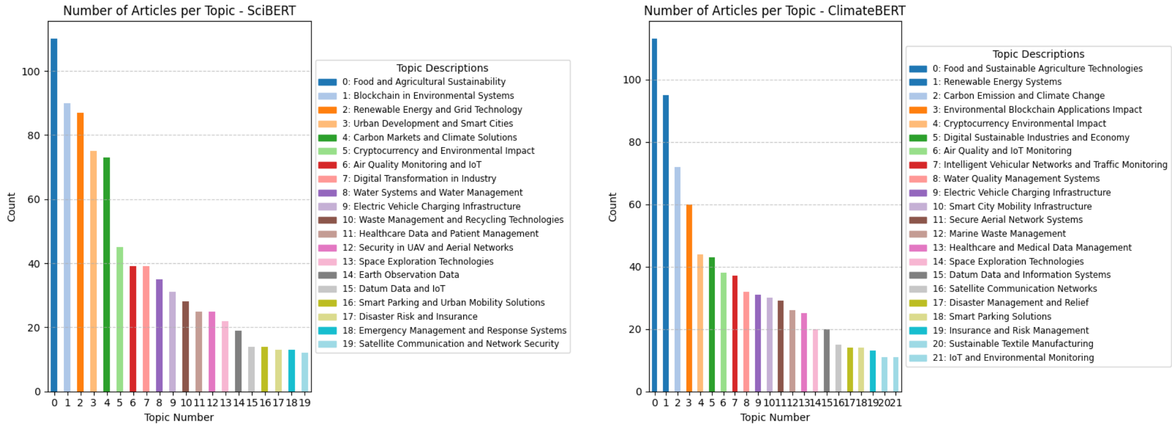
Performing topic analysis with BERT, we report in Table 2 the results obtained. Using SciBERT as an embedding model, we obtained 20 topics; instead, using ClimateBERT, we obtained 22 topics. The barplot of topics obtained by the two methods is shown in Figure 6. The plot of the evolution over time of the 5 topics most commonly found in the literature is shown in Figure 7. For this plot, we considered the time interval from 2017 to 2024 because, as shown in Figure 8, this is the period when there are the most scientific publications on these topics. Finally, the plot of the top 5 most cited topics is shown in Figure 1. For all these three graphs, the results obtained by applying the two different embedding models were compared.

Table 2. Comparison of Topic Extraction with SciBERT and ClimateBERT

Topic extracted with SciBERT		*Topic extracted with ClimateBERT*	
Topic Name	**Count**	**Topic Name**	**Count**
Food and Agricultural Sustainability	116	Food and Sustainable Agriculture Technologies	119
Blockchain in Environmental Systems	171	Renewable Energy Systems	119
Renewable Energy and Grid Technology	115	Carbon Emission and Climate Change	89

Table 2. *Cont.*

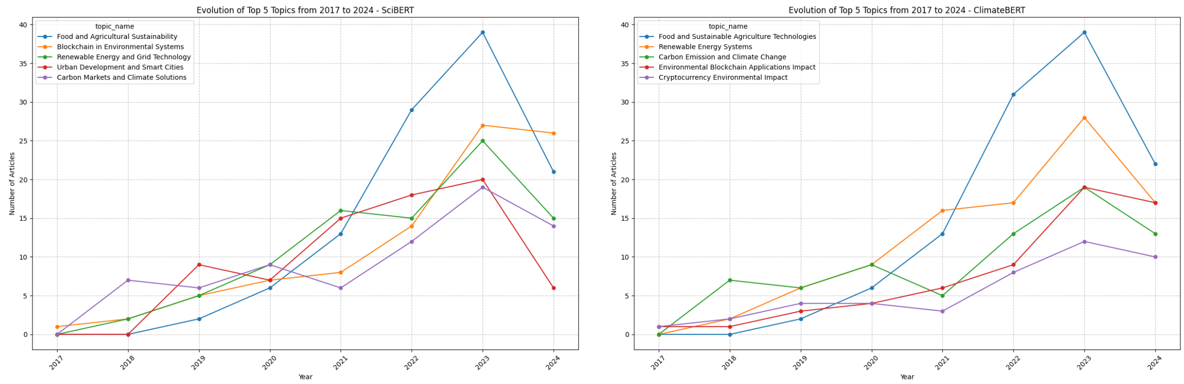
Topic extracted with SciBERT		*Topic extracted with ClimateBERT*	
Topic Name	**Count**	**Topic Name**	**Count**
Urban Development and Smart Cities	100	Environmental Blockchain Applications Impact	140
Carbon Markets and Climate Solutions	87	Cryptocurrency Environmental Impact	52
Cryptocurrency and Environmental Impact	53	Digital Sustainable Industries and Economy	99
Air Quality Monitoring and IoT	47	Air Quality and IoT Monitoring	52
Digital Transformation in Industry	98	Intelligent Vehicular Networks and Traffic Monitoring	48
Water Systems and Water Management	42	Water Quality Management Systems	40
Electric Vehicle Charging Infrastructure	31	Electric Vehicle Charging Infrastructure	34
Waste Management and Recycling Technologies	33	Smart City Mobility Infrastructure	51
Healthcare Data and Patient Management	32	Secure Aerial Network Systems	40
Security in UAV and Aerial Networks	42	Marine Waste Management	31
Space Exploration Technologies	28	Healthcare and Medical Data Management	29
Earth Observation Data	29	Space Exploration Technologies	28
Datum Data and IoT	34	Datum Data and Information Systems	37
Smart Parking and Urban Mobility Solutions	14	Satellite Communication Networks	19
Disaster Risk and Insurance	13	Disaster Management and Relief	19
Emergency Management and Response Systems	17	Smart Parking Solutions	14
Satellite Communication and Network Security	18	Insurance and Risk Management	13
		Sustainable Textile Manufacturing	16
		IoT and Environmental Monitoring	31



(a) Barplot of the topics obtained using SciBERT as embedding model

(b) Barplot of the topics obtained using ClimateBERT as embedding model

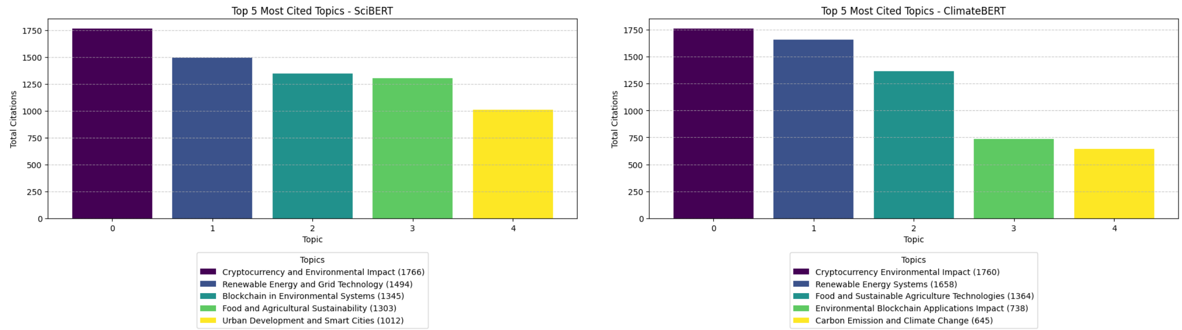
Figure 6. Barplot of the topics obtained by the two topic extraction processes.



(a) Plot of the evolution over time of the top 5 topics obtained with sciBERT

(b) Plot of the evolution over time of the top 5 topics obtained with climateBERT

Figure 7. Plot of the evolution over time of the top 5 topics obtained by the two topic extraction processes.



(a) Top 5 of most cited topics obtained with sciBERT

(b) Top 5 of most cited topics obtained with climateBERT

Figure 8. Top 5 of most cited topics obtained by the two topic extraction processes.

Moving on to the analysis of practical projects extracted from GitHub, we reported in Table 3 the list of topics obtained from the issue analysis by applying the BERTopic model with the BAAI/bge-reranker-bas embedding model. Instead, in Table 4, the topics extracted from the repositories and their clasterisation were reported.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2024 doi:10.20944/preprints202407.1241.v1

Table 3. Issues Topics Extracted with BAAI/bge-reranker-bas

Issues Topic Name	Count
GitHub Repository Management	209
Arxiv Paper Discussions	85
Academic Research and Documentation	84
Scholarly Communication	78
Coding and Development	71
Version Control and Collaboration	56
Software Tools and Configurations	51
Social Media and Personalities	48
Digital Assets and Web Content	43
Development and Issue Tracking	36
Software Documentation	37
Social Media Analysis	35
Web Development	28
Research Publications	27
Technical Configuration	27
Community and Resources	24
Scholarly Research	24
Machine Learning and Research	18
Repository Contributions	19

Table 4. Clusters of Repository Topics

Cluster	Topics
Blockchain Technology Integration	Polkadot Earth Networks, Ethereum for Earth in Asia, Blockchain Planet Projects, Blockchain Messaging Protocols, Blockchain for Earth Preservation, Blockchain Earth Projects, Blockchain Document Management, MES Protocol Ethereum Solana, Blockchain Metaverse Tutorials, Geospatial Blockchain Applications, Decentralized AI Blockchain Token, ERC20 Ethereum Blockchain Token, Blockchain Security and Hacking, Metaverse and Blockchain Essence, IoT and Blockchain Integration, Decentralized Insurance Platform, Global Blockchain Networks, IoT Blockchain Simulations, Mars Currency Blockchain Exchange, Pinball Protocol Blockchain Exchange, Decentralized Blockchain Ledger Oracles
Environmental and Sustainable Projects	Earth-Focused Content Media, Earth-Centric Bitcoin Projects, Hyperledger Earth Projects, Blockchain EarthDAO Ownership, Blockchain for Functional Earths, Blockchain Agriculture Applications, Real Estate Blockchain Fundraising, Blockchain-Powered Christmas Tree, Global Earthcoin Blockchain Village, Smart Security Blockchain Technology, Blockchain Agricultural Sustainability, Environmental Projects and Investments
Cryptocurrency and Financial Transactions	Earth-Centric Bitcoin Projects, Crypto and Ethereum Exchange, Mars Cryptocurrency Blockchain, Kaseicoin Cryptocurrency Platform
Media and Content Creation	Earth-Focused Content Media, Angular Blockchain Explorer, Blockchain Earthcam Image Encryption
Digital Governance and Smart Contracts	Digital DAOs and Jurisdiction, DAO Proposal Management, Blockchain in Government Trust, Blockchain Investment Funds Ledger
Community and Social Impact	Moralis Web3 Metaverse, NFT Assets Blockchain Management, Blockchain and Community Democracy, NFTs Teenagers Platform, Fractal Databases on Facebook
NFT and Digital Assets	Blockchain NFT Earth Projects, NFT Assets Blockchain Management, DAO Proposal Management, NFTs Teenagers Platform

5. Discussion

The interest in blockchain applications in earth, climate, energy, health, etc. has experienced significant growth in recent years, as confirmed by the number of articles in Figure 1. The fact that many of these articles have few citations is due to the fact that many of them are recent, as evidenced by the graph. Nevertheless, the fact that an article related to the energy sector has been so successful testifies to the many benefits provided by blockchain for energy production and buying and selling. One of the main problems found in blockchain application fields is that projects often remain theoretical ideas. Despite this, in that case, the presence of 59 repositories and numerous issues and comments testify to how even developers are working to bring these theoretical ideas to life.

The bibliometric analysis provides us with a first overview of the main topics of discussion in literature. From the full counting analysis, we can observe in Figure 4a that the prevalent topics of interest are related to sustainability, smart city, service, network, energy, emission, waste, product, and market. This is also confirmed by Figure 4c and Figure 4d, which also show the breakdown into clusters, such as agribusiness, green finance and market, waste products, service providers, air pollution and IoT networks, and finally p2p energy trading and electric vehicles. Finally, from Figure 4b, it can be seen that topics such as sustainability overlap across multiple clusters, while others, such as waste, are more specific to their group. On the other hand, from the binary counting analysis, we can analyse the breadth of topics in the dataset and to minimise the influence of prolix terms. Figure 5a shows that we have three big clusters: one related to network, smart city, vehicles, and traffic; another one related to sustainability, innovation, food and agriculture supply chain; and finally, the last one related to energy, carbon emissions, greenhouse gases, and the p2p energy trading market. This result can also be observed in Figure 5c and Figure 5d. Also in this analysis, as can be seen from Figure 5b, topics such as sustainability, supply chain, artificial intelligence, and carbon emissions are cross-cutting across several clusters; in contrast, others such as peer, grid, and driver are more class-specific.

Going into more detail about the topics of interest in scientific articles, in Table 2 we can observe the topic modelling results obtained by applying the two BERT models. The distribution of these is also plotted in Figure 6, providing an even clearer idea of the main interests of the scientific community. In the topics obtained with the SciBERT model, it is noticeable that the most discussed topics concern general topics related to the environment (Blockchain in Environmental Systems), agriculture and food (Food and Agricultural Sustainability), energy (Renewable Energy and Grid Technology), and urban development (Urban Development and Smart Cities). Other smaller topics discuss recycling, emissions, air quality, space and planets, and last but not least, no less interesting aspects of cryptocurrency and mining. These results are also confirmed by the topics obtained with the ClimateBERT model, which in addition extrapolates on topics such as "Sustainable Textile Manufacturing", which is very important nowadays due to the emissions caused by the mass production of clothes and the difficulty of recycling them once they are discarded. The evolution in recent years of these main topics, specifically the top 5, can be seen in Figure 7. Interesting is the steady growth in topics related to agriculture and food, underscoring how even sectors historically characterised by manual labour are facing technological innovations. Also interesting in Figure 7a is the growth in the last year of themes related to the environment in general. This highlights how the cross-cutting use of blockchain for environmental monitoring and prevention is increasingly impactful. One difference between the two models, as noted in Figure 7b, is related to the cryptocurrency topic, which is not present in the top 5 of the SciBERT model. While the latter is the most is the most extreme among the top topics, one related to urban development, which is not highlighted by ClimateBERT instead. This result is interesting in that one would expect more technical topics (cryptocurrencies) to be closer to a model trained on scientific data, while urban development topics are more related to an "environmental" model such as ClimateBERT. Evidence of how cryptocurrencies are impacting the environment, and this aspect is of interest to researchers, is shown in Figure 8, where these topics are the most frequently mentioned. Followed in both cases by topics related to energy, a sector that offers great application possibilities such as

 doi:10.20944/preprints202407.1241.v1

renewable energy certification, energy trading, and energy management. Finally, the other most cited topics concern the environment, agriculture, food, and finally climate change.

Discussing now the results of practical projects on GitHub, let us first observe that Figure 2 and Figure 3 are connected. In fact, we can see that in 2018 and 2022 there was a peak of open repositories, and in 2019 and 2023 there was an increase in open issues probably related to work on new projects by developers.

The topics extracted from the issues and reported in Table 3 show academic and technical topics. Many topics are related to scientific research, Arxiv, etc., highlighting how many scientific researchers probably propose preprints of their solutions to the community to receive feedback and make improvements. Other technical topics concern software tools, digital assets and web content, GitHub repository management, and social media analysis, highlighting the aspects on which blockchain developers in the earth sector most discuss and collaborate. The same results are also confirmed by the analysis of the repositories in Table 4, which probably explain in even more detail some technical aspects such as the use of particular blockchains like Ethereum, Hyperledger, or Polkadot and tokens like ERC-20, NFT, DAO, Crypto, and Bitcoin applications. Despite this, environmental and sustainable project applications are also presented, such as earth-focused projects, blockchain agriculture applications, and real estate blockchain fundraising.

To summarise our results, the first question, **RQ1**, asked: **What are the main discussion topics within scientific literature regarding the use of blockchain in critical areas such as earth sciences, climate change, and environmental health? How have these discussions evolved over time?** Our findings reveal that *food, agriculture, energy, cryptocurrency, carbon emissions, and waste* are the areas of greatest applicability of blockchain, showing steady growth from 2017 to the current time. In these fields, blockchain technology improves *data transparency, security, and collaboration* between the various stakeholders.

The second question, **RQ2**, asked: **What themes emerge from practical blockchain projects in these fields, and how do they develop over time?** The analysis shows that *initial themes revolved around experimental and pilot projects aimed at testing the feasibility of blockchain technologies in real-world contexts, such as transparent tracking of carbon offsets, supply chain management for sustainable resources, and decentralised energy trading platforms*. Another aspect of interest is the use of GitHub to *propose new research works, receive feedback, and make improvements*. The evolution of repositories and issues has *two peaks, one between 2018 and 2019 and one between 2022 and 2023*, demonstrating a growth of interest in recent years also linked to technological innovations.

Finally, the last question, **RQ3**, asked: **Is there an overlap between the themes explored in academic research and those implemented in practical projects?** Considering both analyses, we observe that topics related to *energy, agriculture, and environmental management* are widely discussed in the literature and have practical projects implemented on GitHub. Another aspect in common is the *spike in growth in recent years* towards the use of blockchain in earth sciences, climate change, and environmental health.

6. Threats to Validity

In this study, we recognise a range of potential threats to the validity of the findings, covering external, internal, and construct validity concerns:

Sampling Bias: Our analysis primarily draws from data collected from the Scopus database and GitHub. This sample may not reflect the full diversity of blockchain scientific articles and applications in environmental management. This limitation could influence the extent to which our results can be generalised.

Technological Evolution: The swift advancement in blockchain technology may render our findings less relevant as new platforms or methodologies emerge that were not included in our initial analysis.

Analytical Limitations: The use of bibliometric analysis and BERTopic for topic modelling introduces inherent biases. These methodologies might impose constraints on the data that could overlook subtle or emerging themes. The interpretation of the topics through Chat-GPT 4 and the subsequent check of the authors may also introduce a potential bias due to their background and expertise.

7. Conclusions and Future Works

The study highlights the potential of blockchain technology when aligned with environmental sustainability efforts. The results obtained therefore show the importance of combining theoretical ideas with practical implementations. Using platforms such as GitHub helps researchers get feedback on their proposals and make improvements to the draft projects they implement.

The topics of energy, climate, sustainable mobility, smart cities, food, and sustainable agriculture are becoming increasingly important nowadays. Blockchain fits perfectly with these applications, providing secure, transparent, and immutable records for transactions and data management.

It follows from this that raising the awareness of technicians, researchers, and politicians about the use of this technology in environmental applications is of central importance in society today.

Future research should consider other sources and explore the regulatory, ethical, and security challenges associated with deploying blockchain technology in sensitive environmental and social contexts.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER grant number XXX." and and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

Institutional Review Board Statement: In this section, you should add the Institutional Review Board Statement and approval number, if relevant to your study. You might choose to exclude this statement if the study did not require ethical approval. Please note that the Editorial Office might ask you for further information. Please add "The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of NAME OF INSTITUTE (protocol code XXX and date of approval)." for studies involving humans. OR "The animal study protocol was approved by the Institutional Review Board (or Ethics Committee) of NAME OF INSTITUTE (protocol code XXX and date of approval)." for studies involving animals. OR "Ethical review and approval were waived for this study due to REASON (please provide a detailed justification)." OR "Not applicable" for studies not involving humans or animals.

Informed Consent Statement: Any research article describing a study involving humans should contain this statement. Please add "Informed consent was obtained from all subjects involved in the study." OR "Patient consent was waived due to REASON (please provide a detailed justification)." OR "Not applicable" for studies not involving humans. You might also choose to exclude this statement if the study did not involve humans.

Written informed consent for publication must be obtained from participating patients who can be identified (including by the patients themselves). Please state "Written informed consent has been obtained from the patient(s) to publish this paper" if applicable.

Data Availability Statement: We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section "MDPI Research Data Policies" at https://www.mdpi.com/ethics.

Acknowledgments: This work was partially funded by Ministero dell'Università e della Ricerca (MUR), issue D.M. 351/2022 "Borse di Dottorato"—Dottorato di Ricerca di Interesse Nazionale in "Blockchain e Distributed Ledger Technology", under the National Recovery and Resilience Plan (NRRP).

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflicts of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Bhutta, M.N.M.; Khwaja, A.A.; Nadeem, A.; Ahmad, H.F.; Khan, M.K.; Hanif, M.A.; Song, H.; Alshamari, M.; Cao, Y. A Survey on Blockchain Technology: Evolution, Architecture and Security. *IEEE Access* **2021**, *9*, 61048–61073. doi:10.1109/ACCESS.2021.3072849.
2. Parmentola, A.; Petrillo, A.; Tutore, I.; De Felice, F. Is blockchain able to enhance environmental sustainability? A systematic review and research agenda from the perspective of Sustainable Development Goals (SDGs). *Business Strategy and the Environment* **2022**, *31*, 194–217, https://doi.org/10.1002/bse.2882.
3. Pilkington, M. Blockchain technology: principles and applications. In *Research handbook on digital transformations*; Edward Elgar Publishing, 2016; pp. 225–253.
4. Zhang, A.; Zhong, R.Y.; Farooque, M.; Kang, K.; Venkatesh, V.G. Blockchain-based life cycle assessment: An implementation framework and system architecture. *Resources, Conservation and Recycling* **2020**, *152*, 104512. doi:https://doi.org/10.1016/j.resconrec.2019.104512.
5. Wang, Y.; Chen, H. Blockchain: A potential technology to improve the performance of collaborative emergency management with multi-agent participation. *International Journal of Disaster Risk Reduction* **2022**, *72*, 102867. doi:https://doi.org/10.1016/j.ijdrr.2022.102867.
6. Jin S., C.H. The trends of blockchain in environmental management research: a bibliometric analysis. *Environmental Science and Pollution Research* **2023**, *30*. doi:10.1007/s11356-022-19856-x.
7. O'Donovan, P.; O'Sullivan, D.T.J. A Systematic Analysis of Real-World Energy Blockchain Initiatives. *Future Internet* **2019**, *11*. doi:10.3390/fi11080174.
8. Joshi, P.; Tewari, V.; Kumar, S.; Singh, A. Blockchain technology for sustainable development: a systematic literature review. *Journal of Global Operations and Strategic Sourcing* **2023**, *16*, 683–717.
9. Popkova, E.G.; Bogoviz, A.V.; Lobova, S.V.; Vovchenko, N.G.; Sergi, B.S. Blockchain, sustainability and clean energy transition. *Global Transitions* **2023**, *5*, 64–78. doi:https://doi.org/10.1016/j.glt.2023.04.002.
10. Böckel, A.; Nuzum, A.K.; Weissbrod, I. Blockchain for the Circular Economy: Analysis of the Research-Practice Gap. *Sustainable Production and Consumption* **2021**, *25*, 525–539.
11. Gawusu, S.; Zhang, X.; Ahmed, A.; Jamatutu, S.A.; Miensah, E.D.; Amadu, A.A.; Osei, F.A.J. Renewable energy sources from the perspective of blockchain integration: From theory to application. *Sustainable Energy Technologies and Assessments* **2022**, *52*, 102108. doi:https://doi.org/10.1016/j.seta.2022.102108.
12. Wang, Q.; Su, M. Integrating blockchain technology into the energy sector — from theory of blockchain to research and application of energy blockchain. *Computer Science Review* **2020**, *37*, 100275.
13. Dorfleitner, G.; Muck, F.; Scheckenbach, I. Blockchain applications for climate protection: A global empirical investigation. *Renewable and Sustainable Energy Reviews* **2021**, *149*, 111378.
14. Arshad, A.; Shahzad, F.; Ur Rehman, I.; Sergi, B.S. A systematic literature review of blockchain technology and environmental sustainability: Status quo and future research. *International Review of Economics & Finance* **2023**, *88*, 1602–1622. doi:https://doi.org/10.1016/j.iref.2023.07.044.

15. Kirby, A. Exploratory Bibliometrics: Using VOSviewer as a Preliminary Research Tool. *Publications* **2023**, *11*. doi:10.3390/publications11010010.
16. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* **2022**.
17. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text, 2019,
18. Webersinke, N.; Kraus, M.; Bingler, J.A.; Leippold, M. ClimateBert: A Pretrained Language Model for Climate-Related Text, 2022, arXiv:cs.CL/2110.12010.
19. Röder, M.; Both, A.; Hinneburg, A. Exploring the space of topic coherence measures. Proceedings of the eighth ACM international conference on Web search and data mining, 2015, pp. 399–408.
20. Colavito, G.; Lanubile, F.; Novielli, N.; Quaranta, L. Leveraging GPT-like LLMs to Automate Issue Labeling. 2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR). IEEE, 2024.
21. Vaccargiu, M.; Aufiero, S.; Bartolucci, S.; Neykova, R.; Tonelli, R.; Destefanis, G. Sustainability in Blockchain Development: A BERT-Based Analysis of Ethereum Developer Discussions. Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, 2024, pp. 381–386.