Digital Evolution: blockchain field research

Carlos Bernardino
ISEG – Lisbon School of Economics & Management,
University of Lisbon
Lisbon, Portugal
cmbernardino@phd.iseg.ulisboa.pt

Carlos J. Costa
Advance/CSG, ISEG – Lisbon School of Economics & Management, University of Lisbon
Lisbon, Portugal
cjcosta@iseg.ulisboa.pt

Manuela Aparício
NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa,
Lisbon, Portugal
manuela.aparicio@novaims.unl.pt

Abstract — **Blockchain is a relatively new technology supported by a decentralized database that received special attention at the research level in the last years due to its fundamental characteristics. Numerous researchers have applied blockchain studies in many fields, but the scope is very large and there is no delimitation of new and emergent trends. The method used to better understand the evolution and impact of blockchain technology was a bibliometric analysis. A search was conducted at digital Elsevier's database with a single keyword blockchain, and 23383 articles were collected. The VOSviewer software was used as a tool to construct and visualize bibliometric networks and to build co-occurrences networks. The results indicate that there are positive correlations between countries, that China and the United States are part of the most influential cluster. Indicates also emerging developments in the areas of governance, industry, decision-making processes, management, internet of things, information security, and a new hot topic, energy.**

Keywords - Blockchain; management; trust; bibliometrics; processes; keyword networks, visualization networks.

I. INTRODUCTION

The challenges about industry evolution in the 1980s were amazing, with the evolution of the technology giving the first great step over the internet creation. Since that some relevant achievements stand out, such as Data integration in the 1980s [1], Functionality integration in the 1990s [2]-[3], Services integration in the 2000s [4], Microservices integration in the 2010s [5], and Blockchain integration in the 2020s [6]-[7]. Blockchain evolution can be traced as a result of many technological evolutions, since the 1990s, with the first reference to cryptographically secured blocks [8], then the first decentralized digital currency referenced in 1998 by Nick Szabo, ten years later the first Bitcoin white paper [9] was published, and in 2009 the first block from bitcoin chain was mined. In 2014, Ethereum was forked from the main bitcoin chain and implements smart contracts [10], incorporating some sort of behavior into the technology.

Nowadays, in the modern world the most valuable resource is not money [11], but data, and there are major challenges related to data security and reliability visible not only at the industry level but also at the social level with real implications, for example regarding privacy constraints and misinformation [12]. Blockchain technology seems to solve the security and trust dimensions challenge, bringing new challenges to the industry, and the social, especially for blockchain adoption, as they do not align with the internal processes of the traditional industry. This technology is so innovative and disruptive that it puts pressure on organizations to digitally upgrade, not only in terms of software and hardware but mainly in the way internal processes are designed to be competitive. It has the potential to act as a facilitator for the development and improvement of inter-organizational processes [13] filling the trust gap [13] in business networks [14], and among organizations [16], allowing trust relationships between distrusted partners. Blockchain is supported in five disruptive elements, transparency, immutability, security, consensus, and smart contracts, that altogether state it as a strategic variable to interconnect information technology systems and thus assist organizations in supporting their business processes and creating innovation opportunities across industries [7].

The technology reaches the integration stage in the research and development dimensions, with major regulatory and ethical challenges. In the future, the technology can achieve global integration, where tokenization will take on the relevant role, enabling blockchain implementations in virtually any industry. Currently, digital transformation is enabling the emergence of blockchain, where regulators, technology providers, and new startups combined, can help with the transformative path.

This work aims to identify the state of the art of blockchain technology in the research field. To understand the state of the art on a given topic, several authors have used bibliometric studies to explore the impact of research areas and to better define the object of study [17]–[19]. To improve the understanding of the evolution and impact of blockchain technology, we performed two independent searches with the same keyword «blockchain» in the

digital databases Clarivate Web of Science (WoS) [20] and Elsevier's Scopus database [21] and analyzed selected data using text mining tools and mapping software.

This study contributes to the identification of the countries that most develop theories on blockchain technology, as well as the groups of countries that most influence these developments. This work also presents an analysis of the frequencies of the most used keywords in all the analyzed articles, and their relationship with other closest keywords, indicating directly and effectively which are the research domains, the hottest topics, and the emerging trends.

The theoretical implications of this work are that it can help to understand which the most emerging areas are related to the study and development of blockchain technology in the theoretical field and to delimit future studies.

The practical implications of this work are indicative considerations about the need to research deeply theoretical models related to emerging new trends and related findings.

II. BIBLIOMETRIC RESEARCH

A. *Bibliometric Research: Scopus & Web of Science (WoS)*

The search was conducted at the Scopus database in November 2021 and resulted in 23.383 results, 96% more than the search conducted at the WoS database with 12.697 results, and for that, the bibliometric analysis was conducted at the digital library Scopus over WoS due to greater scientific coverage.

TABLE I. BIBLIOMETRIC RESEARCH: SCOPUS PLUS WEB OF SCIENCE

Digital database	2009-2016	2017	2018	2019	2020	2021	total
Scopus	234	811	2960	5914	7502	5962	23383
web of science	158	591	1902	3488	3434	3124	12697
total	392	1402	4862	9402	10936	9086	36080

a. Note: (Own source).

From the visual analysis to Figure 1, Scopus has more coverage for scientific documentation related to blockchain than the WoS, with greater expression for 2018 to 2021. The exponential growth in both digital databases is revealing of the value, disruption, and importance of the technology.

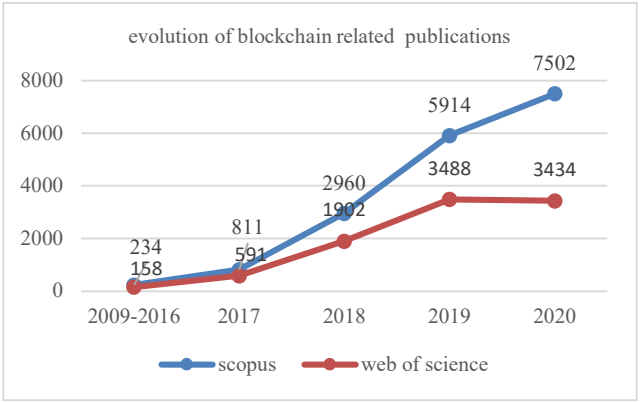

Figure 1 – Publications per year "blockchain "in digital databases (Own source).

III. BIBLIOMETRIC RESEARCH AT SCOPUS

Based on the search in the Elsevier Scopus database we obtained 23,383 results distributed over the years [2009-2016:234], [2017:811], [2018:2,960], [2019:5,914] and [2020:7,502]. The bibliometric data was downloaded with full details, which includes a citation and bibliographical information, abstract & keywords, funding details, and other information. There were limitations to 2000 results for each download on the Scopus front page and to download all the results, several cumulative filters were made, first per year and then by subject, generating 14 files.

After this, all the files were opened simultaneously in the VOSviewer [22] software developed by Eck and Waltman at Leiden University. This software tool is frequently used to construct and visualize bibliometric networks [23] and also to build co-occurrences networks of important terms extracted from a corpus of scientific literature, using text mining functionality [24], [25], that incorporates advanced layout and clustering techniques, and determine the association strength as a normalization measure.

IV. TYPE OF ANALYSIS: CO-AUTHORSHIP – BY COUNTRY

To understand the influence and the impact that a country had on blockchain research, an analysis was performed using the VOSviewer software with the following parameters: Type of analysis – co-authorship; Unit of analysis – country; Counting method – full; Minimum of documents per country – 5; Number of citations per country – 5. This analysis match 97 out of 1073 countries and resulted in indicators about documents, citations, and correlations between them. When analyzing the strength of the links between the documents and citations it's possible to highlight a great correlation between countries. Regarding the 30 most relevant, 3 clusters were generated. In Table II it's possible to detail the clusters individually.

TABLE II. CLUSTER ANALYSIS CO-AUTHORSHIP COUNTRY – TOP30.

cluster	items	links	TLS	Norm. citations	most relevant countries
1	16	390	4074	7781,35	Austria, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Russian Federation, Spain, Sweden, Switzerland, United Kingdom
2	8	182	2553	7057,15	Canada, India, Malaysia, Pakistan, Saudi Arabia, South Korea, Taiwan, United Arab Emirates
3	6	163	5334	15291,46	Australia, China, Hong Kong, Japan, Singapore, United States

Note: (Own source).

An analysis of Table II gives indicators regarding countries' impact and influence on blockchain field research.

- **Cluster 3** with 163 links and 5334 of total link strength it's the cluster with the stronger correlation between countries. The normalized citations of 15291,41 indicate the influence and impact that this cluster has at the level of the research field. It's also interesting to check that the United States, Australia, and China are part of the same strong cluster, naturally with a strong correlation. China and United States are early blockchain researchers that successively support institutionally the development and research of the technology and innovative projects that integrate it.
- **Cluster 1** with 390 links and 4074 of total link strength is composed mostly of European countries, so they have a natural geographic correlation. The partnership between universities and countries is also important to justify it. At the European level, and as an initiative to consolidate Europe's position about this transformative technology, the European Commission created the European Blockchain Observatory and Forum [26].
- **Cluster 2** with 182 links and 2553 of total link strength is composed mostly of countries from Asia, except Canada, so they have a good geographic correlation. It's also the cluster with the weakest correlation between countries, visible by total link strength value.

Despite the major influence that some countries naturally have at the level of scientific research, due to the use of the natural scientific language, the greatest funding research, or even the geographically separation, it's interesting to see that there are different clusters related to countries, and Chinese researchers and USA research cooperate are working closely. Cluster 3 is the cluster with the most influence over the others.

A. Type of Analysis: Co-occurrences – Keywords

An analysis of keyword co-occurrences was made, to assess which areas are developed indicating directly and effectively which are the research domains, the hottest topics, and the emerging trends. Based on all the data collected in the digital library Scopus, the following parameters were defined in the VOSviewer: Type of analysis - co-occurrences; Unit of analysis - all keywords; Counting Method - Full; Minimum occurrences per keyword - 100; Correspondence 268 of 52.213 keywords). The number of keywords to be selected - 268. Normalization method – Association strength. Clusters – 5. The keywords "block-chain", "blockchain", "blockchains", "block chain" and "blockchain technology" were all removed from the visualization since the whole search is centered on itself. To clean the results from duplicated keywords, the occurrences "internet of things", "internet of things (iot)", "iot" and "internet of thing" were aggregated in a single term "internet of things". In the following table, the top30 keywords are listed, ordered by total link strength.

TABLE III. TOP 30 KEYWORDS ORDERED BY TOTAL LINK STRENGTH

nr	keyword	occurrences	TLS
1	internet of things	5325	29633
2	network security	1772	9545
3	digital storage	1634	8637
4	cryptography	1477	7185
5	smart-contract	1695	6942
6	data privacy	1155	6714
7	security	1251	6687
8	ethereum	1342	5398
9	smart contracts	1378	5395
10	information management	947	5178
11	authentication	1050	5080
12	privacy	860	5075
13	bitcoin	1494	4998
14	peer to peer networks	843	4447
15	electronic money	936	4035
16	privacy by design	704	3971
17	access control	724	3924
18	cryptocurrency	1078	3806
19	supply chains	798	3594
20	artificial intelligence	796	3503
21	data sharing	617	3477
22	commerce	789	3465
23	network architecture	567	3401
24	security of data	686	3366
25	big data	701	3356
26	health care	502	3177
27	security and privacy	474	3112
28	cloud computing	605	2918
29	distributed computer systems	490	2664
30	Privacy-preserving	465	2588

Note: (Own source).

The 268 results selected for viewing were grouped into five clusters, according to the relation of the links between keywords and documents. In a more detailed way, each cluster can be seen in the following table, with the description of the words with the most occurrences. This table shows the emergence of new research tendencies regarding the technology, and new use cases that can lead to new internal business processes at the level of design and implementations. Most of the terms listed in **Cluster 1** refer, transversally, to application areas regarding Government and industry, with relevant challenges for business process management (BPM), regarding the way the processes are designed, decision processes, government, industry, management. **Cluster 2** refers clearly to the emergent and hot topic of internet of the thing's research with so many occurrences and great value of total link strength. involving IoT applications; smart cities; privacy and security; vehicle to vehicle communications. **Cluster 3** refers mainly to internal aspects of blockchain technology development and applications regarding operations and management decisions. **Cluster 4** refers to applications regarding assurance and data integrity. **Cluster 5** refers to an innovative application area regarding blockchain technology, the energy area. In table 4 we show relevant clusters of keywords related to the blockchain research development, indicating as the initial objective state the hottest topics and emergent trends related to blockchain. Table IV highlight Cluster 1 the most relevant keywords grouper by cluster, and in

TABLE IV. MOST RELEVANT KEYWORDS GROUPED BY CLUSTER

cluster	items	color	links	most relevant keywords	application area
cluster 1	98		286	distributed ledger; smart contracts; information systems; supply chain; information management; decision making; copyrights; computer science; cryptocurrencies; crowdsourcing decision making; digital economy; economic and social effects; electronic voting; Ethereum; finance; fintech; food supply; identity management; Hyperledger; insurance; industry 4.0; laws and legislation; proof of concept; real-world; robotics; social networking; supply chain; sustainability; technology adoption; traceability; transparency; trust	Industry Government Management Decision Processes Business Process Management
cluster 2	58		292	Identity authentication; 5g mobile communications; automation; trust management; cyber security; data acquisition; denial-of-service; trusted computing; internet of things; intelligent buildings; smart homes; IoT applications; smart cities; privacy and security; vehicle to vehicle communications	IoT
cluster 3	48		264	complex networks; computation theory, consensus algorithm; consensus protocols; cryptocurrency; deep learning; electronic money; learning algorithms; miners; mining; permissioned blockchain; scalability; game theory; performance analysis bitcoin; distributed computer systems; distributed networks; electronic money, electronic health record; medical computing; technology; medical record; diagnosis; privacy	Security of software
cluster 4	39		273	cryptocurrency, data privacy; data share; data security; data integrity; distributed storage; cryptography; security of data; trusted third parties, electronic document exchange; electronic health records; distributed storage	Assurance Data integrity
cluster 5	20		291	peer to peer networks, electric power transmission, power markets, smart power grids; decentralized networks energy; energy trading; microgrids	Energy

Note: The colors of each item in Table IV correspond to the colors of figure 2. (Own source).

To facilitate the data interpretation of the network maps, it's important to mention some concepts used by the VOSviewer software, for example in network visualization items are shown by a circle with a label, the volume of the circle and the size of the label depend on the item's importance, the color of each item is related to the cluster assigned to a group of items, the lines between items represent links, the distance between items is also a factor and the closer the item is, the stronger its relationship [24], [25]. The colors are helping to indicate directly and effectively the research domains, the hottest topics, and the emerging trends.

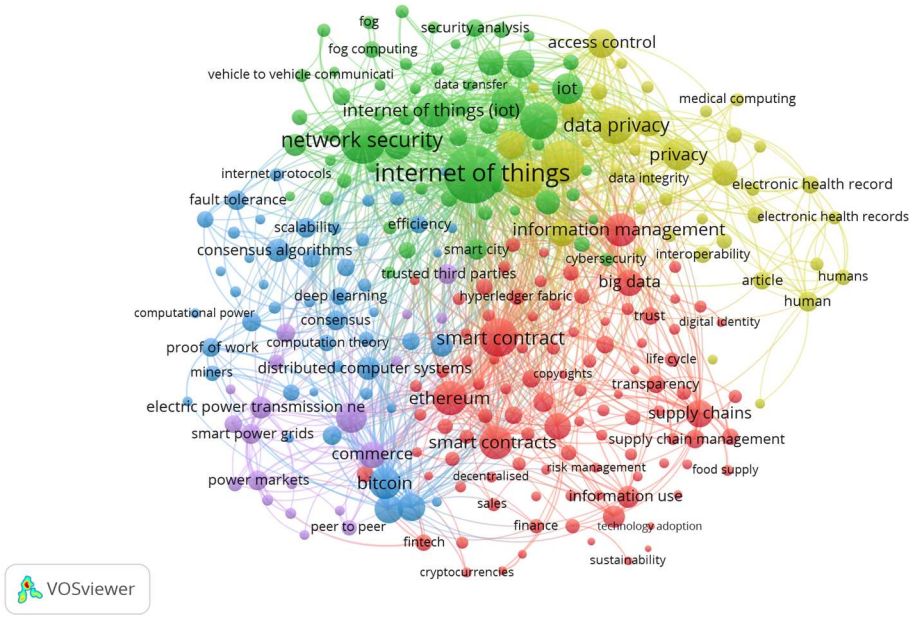

Figure 2 – Network Visualization of Co-occurrence Clusters (Own source).

By the distribution of the occurrences on the map, in Figure 3 it is possible to visualize proximity relations and through the lines (links) that connect them and visualize the intensity of each relation.

- **Cluster 1**. Stands out for its position close to cluster 4 and cluster 5, which is a close relationship between the terms: smart contracts; information systems; supply chain; information management; decision making; electronic money; proof of work; distributed computer systems, cryptocurrency, consensus protocol e data privacy; cryptography; block-chain; security of data; trusted third parties. Indicate trends related to BPM, decision processes, industry, Government, and industry with assurance, data integrity, and the energy sector.

- **Cluster 2**. Closer to clusters 4, so list the terms: human; electronic health record; medical computing; technology; medical record; diagnosis; privacy; peer-to-peer networks; commerce, electric power transmission; smart power grids; power markets e smart contracts; information systems; supply chain; information management and decision making. Indicate trends related to the security of software, data integrity, and assurance.

- **Cluster 3.** Close relationship with clusters 1 and 5 across smart contracts; information systems; supply chain; information management; decision-making and data privacy; cryptography; blockchain; security of data and trusted third parties. Indicate trends related to BPM, decision processes, Government and industry with assurance, data integrity, and the energy sector.

- **Clusters 4 and 5.** Close relationship with clusters 1 and 2, establishing links between the terms smart contracts; information systems; supply chain; information management; decision-making e internet of things; intelligent buildings; smart homes; IoT applications and smart cities. This cluster indicates trends related to assurance, data integrity, and the energy sector, and this is a new hot and emerging topic.

To better understand the framework of some topics related to management, on focus in Cluster 1, and the challenges related to the alignment of the business processes to the blockchain technology, the composed term «business process» was highlighted. The number of occurrences for the keywords "business process " has only 132 occurrences and 479 total link strength. It is an indicator that there are few studies in these specific and critical areas, regarding the adoption and integration of blockchain technology.

V. CONCLUSIONS

The main objective of this work is to achieve directly and effectively which are the main research domains, the hottest topics, and the emerging trends regarding Blockchain technology. The method used was a bibliometrics search at digital databases. We performed two independent searches with the same keyword «blockchain» in the digital databases Clarivate Web of Science and Elsevier's Scopus database and analyzed selected data using text mining tools and mapping software to improve the understanding of the evolution and impact of blockchain technology delimiting the frontiers of the field.

Some results show that despite the major influence that some countries naturally have at the level of scientific research, due to the use of the natural scientific language, or the greatest funding research, there are correlations between countries and it's clear that a cluster led by China and the United States who are early blockchain researchers with successively institutionally and funding support are the most influential researchers. The results show also that there is a cluster mainly composed of the European countries, indicating a strong natural geographic correlation. The results also show the early trends and frontiers of blockchain research and related to first to industry, management, government, and decision processes; the second is very clearly and deeply related to the internet-of-things; the third is related to the security of software; the fourth is related to assurance and data integrity, and the last trend related to energy. Regarding energy, this is a new hot topic that is emerging. All this implies that Blockchain is not only related to cryptocurrencies but is also part of a whole new technological revolution with emerging developments in the areas of governance, industry, decision-making processes, and various areas of the internet of things, information security, and energy.

The theoretical implications of this work are that it can help to understand which the most emerging areas are related to the study and development of blockchain technology in the theoretical field and to delimit future studies.

The practical implications of this work are indicative considerations about the need to research deeply theoretical models related to emerging new trends and related findings.

The search was not conducted in all existing digital databases, only Scopus and WoS, which is understood to be a limitation. Another limitation may be related to the classification of the articles' subjects in the digital databases and their area of development not being correlated to the analysis based on text mining as performed using the VOSviewer tool in this work.

For future work, it would be important to repeat the research carried out in the same digital library, and the analysis to assess differences in the definition of boundaries related to the blockchain emerging themes and hot topics, always to better delimit the field.

REFERENCES

[1] S. Narayanan, A. S. Marucheck, and R. B. Handfield, 'Electronic Data Interchange: Research Review and Future Directions*', *Decis. Sci.*, vol. 40, no. 1, pp. 121–163, 2009, doi: 10.1111/j.1540-5915.2008.00218.x.

[2] D. C. Schmidt, D. L. Levine, and S. Mungee, 'The Design of the TAO Real-Time Object Request Broker', *Comput. Commun.*, vol. 21, pp. 294–324, 1997.

[3] J. C. Soo Meng, 'Donald Schön, Herbert Simon and The Sciences of the Artificial', *Des. Stud.*, vol. 30, no. 1, pp. 60–68, Jan. 2009, doi: 10.1016/j.destud.2008.09.001.

[4] M.-T. Schmidt, B. Hutchison, P. Lambros, and R. Phippen, 'The Enterprise Service Bus: Making service-oriented architecture real', *IBM Syst. J.*, vol. 44, no. 4, pp. 781–797, 2005, doi: 10.1147/sj.444.0781.

[5] R. Battle and E. Benson, 'Bridging the semantic Web and Web 2.0 with Representational State Transfer (REST)', *J. Web Semant.*, vol. 6, no. 1, pp. 61–69, fevereiro 2008, doi: 10.1016/j.websem.2007.11.002.

[6] M. Xu, X. Chen, and G. Kou, 'A systematic review of blockchain', *Financ. Innov.*, vol. 5, no. 1, p. 27, Jul. 2019, doi: 10.1186/s40854-019-0147-z.

[7] J. S. Arun, J. Cuomo, and N. Gaur, *Blockchain for business*. Addison-Wesley Professional, 2019.

[8] S. Haber and W. S. Stornetta, 'How to Time-Stamp a Digital Document', in *Advances in Cryptology-CRYPTO' 90*, Berlin, Heidelberg, 1991, pp. 437–455. doi: 10.1007/3-540-38424-3_32.

[9] S. Nakamoto, 'Bitcoin: A Peer-to-Peer Electronic Cash System', *2017 Blockchain Technol. Digit. Curr. Natl. Inst.*, p. A-1, 2017.

[10] V. Buterin, 'Ethereum white paper', *GitHub Repos.*, vol. 1, pp. 22–23, 2013.

[11] 'The world's most valuable resource is no longer oil, but data', *The Economist*, May 06, 2017. Accessed: Feb. 18, 2022. [Online]. Available: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

[12] S. Aral, *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health--And How We Must Adapt*. Currency, 2020. [Online]. Available: https://books.google.pt/books?id=N1v6DwAAQBAJ

[13] V. A. de Sousa and B. Corentin, 'Towards an integrated methodology for the development of blockchain-based solutions supporting cross-organizational processes', in *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, May 2019, pp. 1–6. doi: 10.1109/RCIS.2019.8877045.

[14] M. Deutsch, 'Trust and suspicion', *J. Confl. Resolut.*, vol. 2, no. 4, pp. 265–279, 1958, doi: 10.1177/002200275800200401.

[15] R. M. Morgan and S. D. Hunt, 'The Commitment-Trust Theory of Relationship Marketing', *J. Mark.*, vol. 58, no. 3, p. 20, Jul. 1994, doi: 10.1177/002224299405800302.

[16] B. Carminati, E. Ferrari, and C. Rondanini, 'Blockchain as a Platform for Secure Inter-Organizational Business Processes', in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, Oct. 2018, pp. 122–129. doi: 10.1109/CIC.2018.00027.
[17] M. Aparicio, F. Bacao, and T. Oliveira, 'An e-Learning Theoretical Framework', *J. Educ. Technol. Soc.*, vol. 19, no. 1, pp. 292–307, 2016.
[18] F. Bento, C. J. Costa, and M. Aparicio, 'ERP Conceptual Ecology', in *New Knowledge in Information Systems and Technologies*, Cham, 2019, pp. 351–360. doi: 10.1007/978-3-030-16181-1_33.
[19] D. Silva, M. Aparicio, and C. J. Costa, 'Free and open source bibliometric study', in *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, Jun. 2019, pp. 1–6. doi: 10.23919/CISTI.2019.8760695.
[20] 'Web of Science', May 20, 2021. http://apps.webofknowledge.com/ (accessed May 20, 2021).
[21] 'Scopus', May 2021. https://www.scopus.com/ (accessed May 20, 2021).
[22] N. Eck and L. Waltman, *VOSviewer*. 2019.
[23] Y. Jin and S. Ji, 'Mapping hotspots and emerging trends of business model innovation under networking in Internet of Things', *EURASIP J. Wirel. Commun. Netw.*, vol. 2018, no. 1, p. 96, maio 2018, doi: 10.1186/s13638-018-1115-4.
[24] J. A. Moral-Munoz, A. G. López-Herrera, E. Herrera-Viedma, and M. J. Cobo, 'Science Mapping Analysis Software Tools: A Review', in *Springer Handbook of Science and Technology Indicators*, W. Glänzel, H. F. Moed, U. Schmoch, and M. Thelwall, Eds. Cham: Springer International Publishing, 2019, pp. 159–185. doi: 10.1007/978-3-030-02511-3_7.
[25] N. J. van Eck and L. Waltman, 'Citation-based clustering of publications using CitNetExplorer and VOSviewer', *Scientometrics*, vol. 111, no. 2, pp. 1053–1070, May 2017, doi: 10.1007/s11192-017-2300-7.
[26] 'EU Blockchain Observatory & Forum', *EUBlockchain*. https://www.eublockchainforum.eu/eu-blockchain-observatory-forum (accessed Feb. 10, 2022).