

Article

Trend Analysis of Decentralized Autonomous Organization Using Big Data Analytics

Hyejin Park ¹, Ivan Ureta ² and Boyoung Kim ^{1,*} 
¹ Seoul Business School, aSSIST University, Seoul 03767, Republic of Korea; hjpark@assist.ac.kr

² Department of Business Economics, Health and Social Care, The University of Applied Sciences and Arts of Southern Switzerland, 6928 Manno, Switzerland; ivan.uretavaquero@supsi.ch

* Correspondence: bykim2@assist.ac.kr; Tel.: +82-10-4046-2428

Abstract: Decentralized Autonomous Organizations (DAOs) have gained widespread attention in academia and industry as potential future models for decentralized governance and organization. In order to understand the trends and future potential of this rapidly growing technology, it is crucial to conduct research in the field. This research aims at a data-driven approach for the objective content analysis of big data related to DAOs, using text mining and Latent Dirichlet Allocation (LDA)-based topic modeling. The study analyzed tweets with the hashtag #DAO and all Reddit data with “DAO”. The results were from the identification of the top 100 frequently appearing keywords, as well as the top 20 keywords with high network centrality, and key topics related to finance, gaming, and fundraising, from both Twitter and Reddit. The analysis revealed twelve topics from Twitter and eight topics from Reddit, with the term “community” frequently appearing across many of these topics. The findings provide valuable insights into the current trend and future potential of DAOs, and should be used by researchers to guide further research in the field and by decision makers to explore innovative ways to govern the organizations.

Keywords: decentralized autonomous organization (DAO); blockchain; big data analytics; text mining; network analysis



Citation: Park, H.; Ureta, I.; Kim, B. Trend Analysis of Decentralized Autonomous Organization Using Big Data Analytics. *Information* **2023**, *14*, 326. <https://doi.org/10.3390/info14060326>

Academic Editor: Leonardo Nascimento Ferreira

Received: 17 May 2023

Revised: 6 June 2023

Accepted: 8 June 2023

Published: 9 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Decentralized Autonomous Organization (DAO) has emerged as a popular blockchain-based application, characterized by organizations built on smart contracts capable of autonomous execution [1]. With over 10,000 DAOs established worldwide since their introduction in April 2016, the funds deposited in these DAOs have grown significantly, surpassing USD 23B recently [2]. A distinguishing feature of DAOs is their lack of central control, as their rules are encoded in blockchain-based smart contracts, enabling autonomous and automated operations on a distributed peer-to-peer network. This decentralized mechanism enhances visibility, traceability, accountability, and transparency within the organization, mitigating the costs associated with human mistrust and unfulfilled promises, and improving decision-making efficiency and transparency.

Despite the failure of the initial DAO, The DAO, in April 2016, resulting from a security breach caused by a vulnerability in its smart contract, subsequent DAO projects have continued to emerge and flourish. Various entities, including original blockchain initiatives such as NFTs and DeFi as well as traditional organizations such as corporations, venture capital firms, and political parties, have embraced DAOs as a novel approach to constructing and operating organizations, attracting members, and managing funds. However, despite the growing interest in DAOs and their potential for mass adoption, limited data-driven research exists to comprehend the overall trends and future possibilities associated with DAOs.

Previous studies have primarily focused on case analyses, particularly examining The DAO as the first DAO established on the Ethereum platform [3–5]. Additionally, they have

explored the historical understanding, use cases, and societal implications of DAOs [6–8], as well as the adoption of DAOs and their implications for existing governance systems and businesses [9–15]. The existing studies have investigated the underlying technology of DAOs, including their advantages, limitations, and potential solutions [16].

These endeavors have all been made in order to understand and utilize the emerging technology and tools represented by DAOs. In recent times, an increasing number of studies have contributed to the academic discourse on DAOs. For instance, Lu Liu et al. [8] comprehensively classified the latest research that combines blockchain and DAO, presenting related studies in various domains, evaluating recent advancements, and predicting future development directions such as corporate governance, blockchain government, and social DAOs for crowdfunding.

By examining the trends and limitations of previous studies, this paper aims to address this research gap by employing data-driven analysis using text-mining technology to investigate the current state of DAOs and propose future research directions and practical applications. While Twitter has been widely utilized as a source of information for blockchain and cryptocurrency research, with a predominant focus on cryptocurrency prices [17–20], this study extends beyond price-related investigations by analyzing Twitter data specifically pertaining to DAOs. Furthermore, this research capitalizes on another valuable data source, Reddit, which is acknowledged for its ability to provide insights into general trends within rapidly growing industries [21].

This research identifies key topics, keywords, and trends associated with DAOs, such as NFTs, finance, gaming, and fundraising. The recurring presence of the term “community” will also be explored to understand its significance in the context of DAOs. The contributions of this study extend to academia, industry, and institutional stakeholders, providing insights that facilitate informed discussions and contribute to future research and practical applications across diverse domains in the field of DAOs.

Furthermore, the implications of this study extend to legal and ethical considerations surrounding DAOs, shedding light on the need for robust frameworks and regulations that address potential challenges and ensure responsible and transparent operations. By addressing these key aspects, this research contributes to the development of a sustainable and inclusive ecosystem for DAOs, unlocking their full potential for the benefit of stakeholders and society as a whole.

2. Methods

2.1. Research Design

This research presents a data-driven methodology utilizing text mining techniques, as illustrated in Figure 1. The methodology comprises three core steps: Data selection and aggregation, Data pre-processing, and Data analysis. The data aggregation process involves two phases: data fetching and language detection. In the data fetching phase, the Rust-based fetcher searched for all tweets containing the hashtag “#DAO” since the inception of DAOs in April 2016. To identify the trend and main topics surrounding DAO, this study aggregated all the English-language tweets containing the hashtag “#DAO”, as hashtags are used on Twitter to express the main topic or the most important word of a short message.

By aggregating tweets with this specific hashtag, we were able to focus our analysis on relevant tweets that were specifically related to DAO and avoid including irrelevant tweets where the term “DAO” was used for a different purpose or meaning. The use of the hashtag ‘#DAO’ ensured that our analysis was focused on the specific topic of interest and provided a clearer picture of the trend and main topics being discussed. Through this selection, it generated 3,885,266 tweets in total. Given the inherent characteristics of DAOs and Web3, where users frequently possess experience, knowledge, influence, and ownership simultaneously, this research considered all texts that deliberately referenced DAOs through the use of hashtags, including user mentions (see Figure 2).

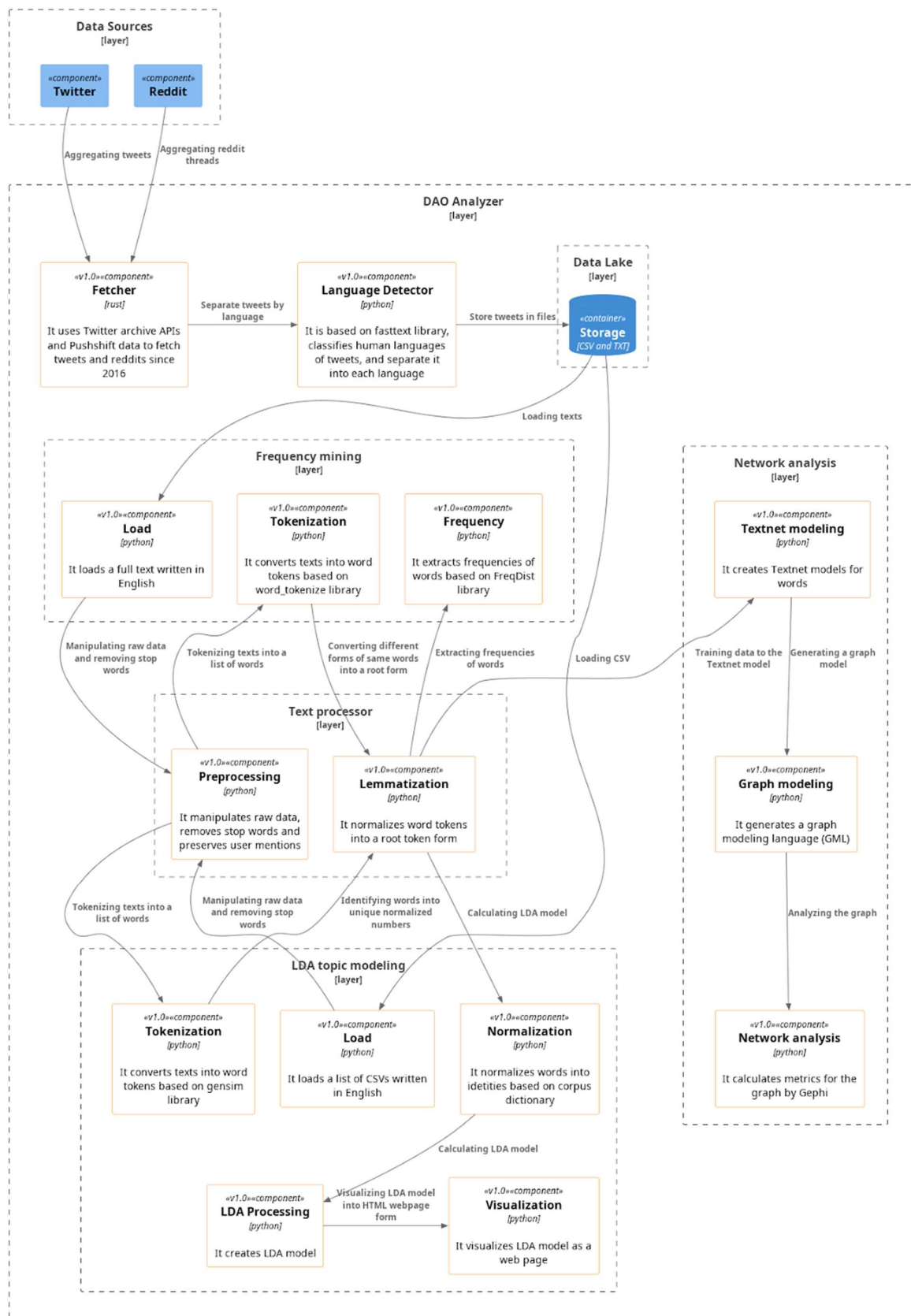


Figure 1. Overall architecture of the research.

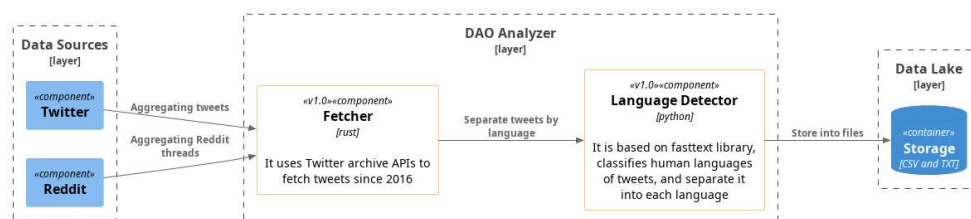


Figure 2. The process of data aggregation.

Notably, the number of tweets related to DAOs in 2022 was four times higher than the total number of tweets from 2016 to 2021 (see Figure 3, Table 1). However, we did not incorporate the time factor into our research analysis. Additionally, because the fetcher aggregates tweets in over 20 languages, the tweets should be classified into their respective languages in the following phase.

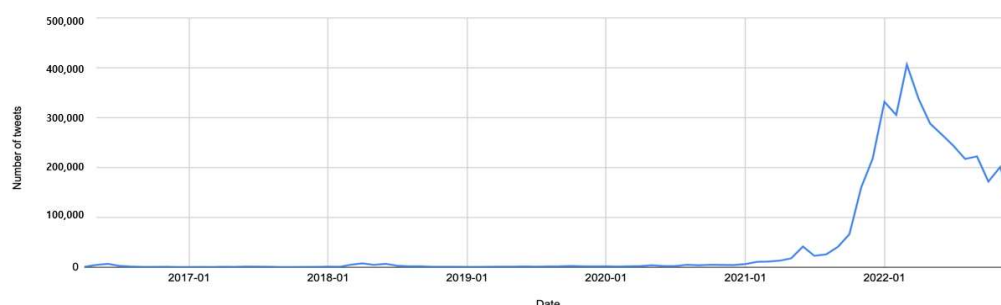


Figure 3. Monthly distribution of tweets related to DAO (accessed on 12 March 2023).

Table 1. Monthly number of tweets related to DAO.

Month	2016	2017	2018	2019	2020	2021	2022
1	0	560	1323	874	1935	6447	331,889
2	0	816	1046	806	1547	10,930	305,613
3	0	461	4967	1005	1858	11,397	406,434
4	1070	982	7780	1279	2483	13,197	337,362
5	4590	699	4746	1240	4045	17,856	288,226
6	6998	1414	6912	1606	2399	41,562	266,064
7	2855	1331	3088	1296	2441	23,031	243,856
8	1365	938	1982	1767	5112	25,871	217,281
9	798	524	1834	1886	3976	41,101	222,478
10	668	431	906	2681	5040	66,289	171,971
11	1002	674	1089	1949	4666	160,238	201,046
12	552	649	1047	1671	4437	217,572	133,459
Annual total number of tweets							
	19,898	9479	36,720	18,060	39,939	635,491	3,125,679
Total number of tweets 3,885,266							

For Reddit, we initially retrieved all submissions from the Pushshift archiving directory. (<https://files.pushshift.io/reddit/submissions/>) (accessed on 12 March 2023). Similar to Twitter's hashtag, Reddit uses the term subreddit. Although a subreddit for DAO exists, it has a limited number of submissions. Therefore, we aggregated all submissions from the DAO subreddit and those that contained 'DAO' in their title (see Figure 4, Table 2).

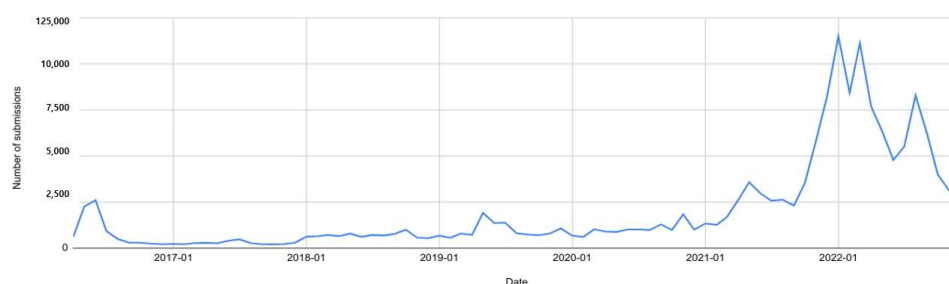


Figure 4. Monthly distribution of Reddit submissions related to DAO (accessed on 12 March 2023).

Table 2. Monthly number of Reddit submissions related to DAO.

Month	2016	2017	2018	2019	2020	2021	2022
1	0	232	624	674	677	1340	11,505
2	0	214	646	568	610	1263	8426
3	0	279	716	799	1020	1692	11,129
4	620	293	653	721	904	2615	7696
5	2253	265	791	1913	880	3582	6356
6	2603	408	613	1365	1011	2972	4782
7	927	481	717	1383	1013	2577	5517
8	502	273	688	809	984	2633	8288
9	296	211	775	746	1292	2316	6254
10	298	207	1002	704	988	3531	3995
11	246	224	572	794	1835	5863	3133
12	212	296	545	1074	1003	8245	3337
Annual total number of reddit submissions							
	7957	3383	8342	11,550	12,217	38,629	80,418
Total number of submissions 162,496							

In the language detection phase, the fasttext library in Python was used to detect the human language of each data. The tweets and Reddit submissions were recorded as a CSV file containing their creation time, which was grouped by month. The data were then accumulated monthly and written in a separate TXT file for each month. The data aggregation process generated monthly CSV and TXT files, where each CSV file contained the specific dates and content of tweets and submissions, and each TXT file contained only the contents of them, which were used for analysis. This process provided a consistent and structured dataset for the subsequent analysis.

The purpose of pre-processing in this study was to prepare the aggregated texts for word frequency analysis, network analysis, and LDA topic modeling. The pre-processing procedure consisted of two stages: removing unnecessary letters and symbols, and eliminating stop words. In the first stage, the text was cleaned by removing phrases indicating URLs and symbols other than alphabetic and numeric characters. The next step involved removing stop words in English using the NLTK library in Python. The NLTK library contains approximately 40 English prepositions, and other words such as “DAO” and “DAOs” were added for analysis.

Typically, text mining would include WordNet lemmatization, but it was omitted from the proposed approach. Instead, the lemmatization integrated various forms of blockchain terminologies into a single form. For instance, coin symbols such as ETH, and BTC were integrated into the platform names Ethereum and Bitcoin, respectively (Table 3).

Table 3. Word integration through lemmatization.

Result Word	Origin Words	Description
Regular expressions	[^a-zA-Z0-9]	Removes all letters excluding alphabets, numbers and space
	http\S+	Removes all URLs led by http
	https\S+	and https
ethereum	eth, ether	Integrates coin symbols into their platform name
solana	sol	
bitcoin	btc	
binance	bsc	
smartcontract	smart contract, smart contracts, smartcontracts	Integrates several terminologies into a single form
dapp	dapps	
reward	rewards	
token	tokens	
nft	nfts	
coin	coins	
project	projects	

Additionally, several terms in the blockchain industry were standardized into a single form. The reason for omitting WordNet lemmatization is two-fold. Firstly, it has a minimal impact since the research focuses on technical keywords rather than verbs or common nouns to analyze trends. The results were almost identical with or without WordNet lemmatization applied as pre-processing. Secondly, technically significant expressions such as “web3” may be compromised in converting different word expressions into root expressions via WordNet lemmatization. As a result, WordNet lemmatization was not implemented in this study.

The extended stop words, which we defined to filter out irrelevant words, are presented in Table 4 below. These stop words primarily consist of prepositions and adjectives. Notably, the predefined keywords “amp” and “rt” were also included in the list of stop words. Additionally, we removed the terms “dao” and “daos” from the list of stop words. This exclusion was because these words represent conceptual keywords, which we aimed to analyze, rather than terms for analysis. In the first stage of experimentation, we eliminated these stop words to improve the accuracy of our analysis.

Table 4. The extended stop words.

1	amp	2	dao	3	daos	4	based
5	rt	6	us	7	one	8	theres
9	via	10	great	11	good	12	done
13	back	14	get	15	best	16	dont
17	anywhere	18	today	19	like	20	Time
21	hello	22	im	23	retweet		

2.2. Data Analysis

Word frequency analysis is a crucial technique that explored the frequency of words appearing in tweets with ‘#DAO’ or Reddit submissions containing ‘DAO’ in their title from April 2016 to December 2022. This method focuses on analyzing contents without considering their creation date. The frequency analysis process is composed of five distinct steps, which are data loading, data pre-processing, data tokenization, lemmatization, and frequency analysis (see Figure 5).

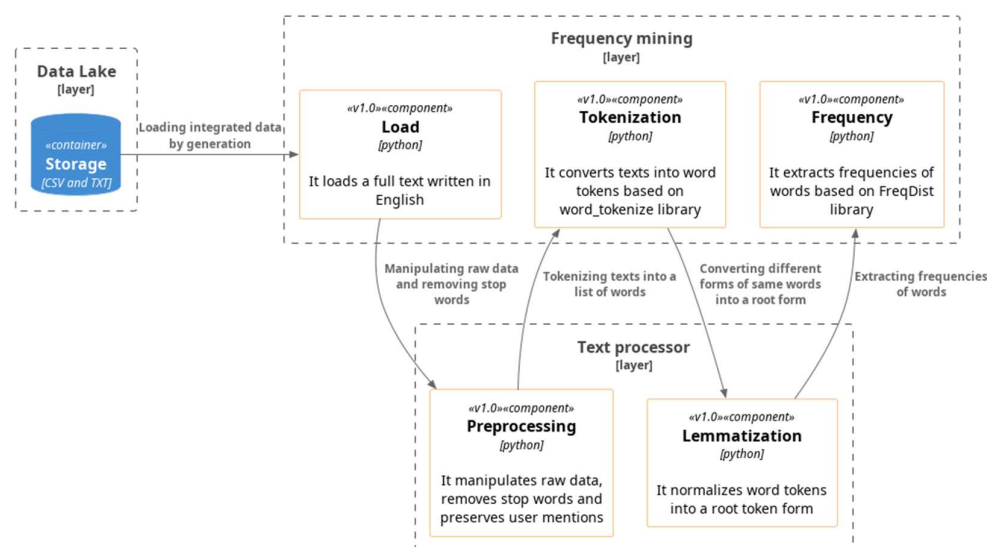


Figure 5. The process of frequency analysis.

The data loading phase starts by reading all the TXT files and loading their contents as a single text into memory. At this stage, the data loading phase was limited to English-language tweets, as the focus of this study was specifically on analyzing English-language content. This text was then subjected to the pre-processing phase, as described in Section 2.2, where the text is cleansed of unnecessary information such as URLs and symbols. The next step, tokenization, involves converting the text into a list of word tokens. This is carried out by treating all space-separated letters as individual tokens. For example, the sentence “This research focuses on trend analysis of DAO” would be tokenized into “This”, “research”, “focuses”, “on”, “trend”, “analysis”, “of”, “DAO”.

The lemmatization phase involves removing stop words and converting technical keywords into a root form. First, the stop words are defined using the NLTK library, which contains approximately 40 English prepositions, along with other extended stop words defined by us. The extended stop words include “amp”, “dao”, “daos”, “rt”, “us”, “one”, “via”, “great”, “good”, “back”, “get”, “best”, “based”, “today”, “like”, “theres”, “dont”, “anywhere”, “done”, “time”, “hello”, “im”, and “retweet”.

The second step of lemmatization involves converting technical keywords into a root form, as discussed in Section 2.2. The final phase involves calculating the frequency of occurrences of words in the token list and identifying the 100 most frequent words. In conclusion, word frequency analysis provides a valuable insight into the frequency of words appearing in contents and is an essential step in text analysis. The five-step process of data loading, pre-processing, tokenization, lemmatization, and frequency analysis ensures that the results are accurate and relevant.

This section details the seven-step LDA topic modeling-based analysis process, which includes data loading, pre-processing, tokenization, lemmatization, normalization, LDA processing, and visualization. During the data storage process, every tweet and Reddit thread was saved as an individual cell in the data lake in the form of a CSV or TXT file. Specifically, each tweet and Reddit thread was saved as a separate line in the TXT file. In the data loading phase, tweets and Reddit threads were loaded from the CSV file and the list of tweets and threads were loaded into memory. The subsequent pre-processing, tokenization and lemmatization phases were the same as those described for frequency analysis (see Figure 6).

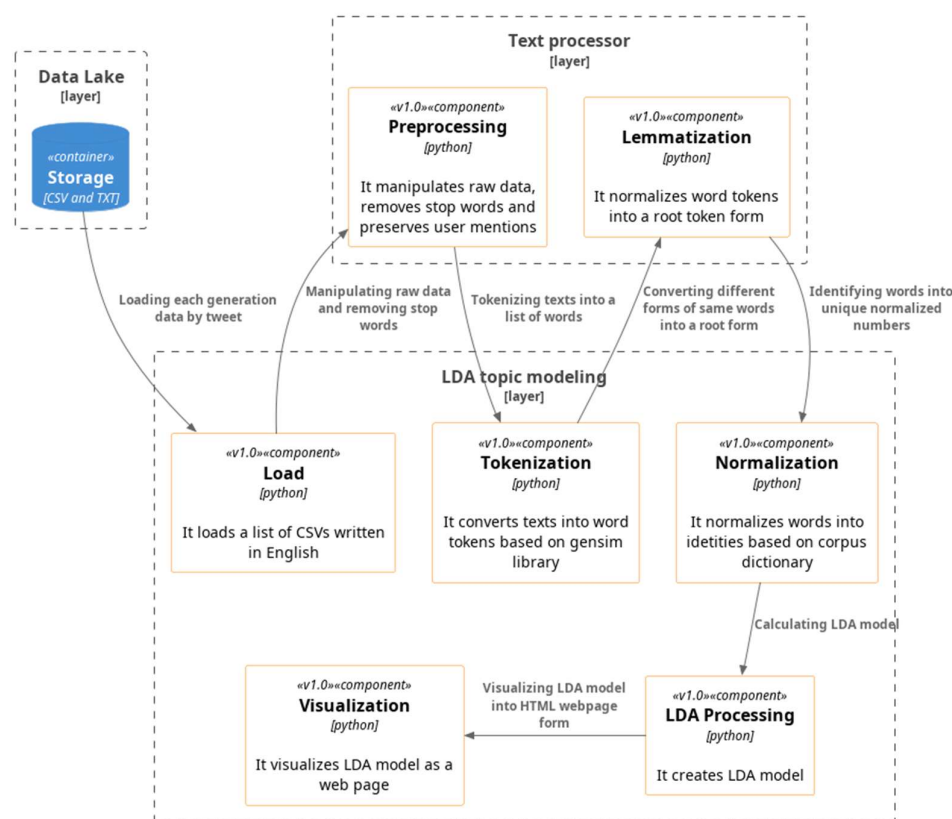


Figure 6. The process of LDA topic modeling.

The normalization phase involved converting the tokenized words into normalized IDs based on the Corpora dictionary of the Gensim library in Python. This normalization process converts the word tokens formed as strings into numerical forms that unify their meanings. The LDA processing phase generates an LDA model using the Gensim library, and the resulting model is visualized in a webpage format using the pyLDAvis library. This visualization allows for more accessible analysis of the results of the LDA analysis.

Latent Dirichlet Allocation (LDA) is founded on the concept that documents comprise a set of prospective topics, and each document is composed of words assigned with probability values that indicate their relevance to a certain number of topics, k [22]. To determine the optimal number of topics for each platform, a range of 5 to 20 topics was generated and their coherence was evaluated using the C_V measure. The C_V score provides a measure of topic coherence, with values closer to 1 indicating higher coherence. However, it is important to acknowledge that the selection of the number of topics is not solely determined by the C_V score, but also considers the researchers' domain knowledge and expertise. In this study, the number of topics was determined by focusing on the most salient and meaningful topics among the highest C_V scores. Based on this consideration, 12 topics were chosen for Twitter while 8 topics were selected for Reddit (see Table 5).

Network analysis is a technique that utilizes text mining to identify the relationship between extracted keywords, thereby enabling the identification of the structure within a network [23]. Centrality analysis is another approach that involves extracting sentence-level text from a document, by dividing it into keywords and analyzing the connections between the extracted keywords to understand the meaning and structure [24]. By using keywords as actors and connections between actors as nodes, it becomes possible to analyze current situations and trends. Centrality analysis includes degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality [25].

Table 5. C_V coherence scores.

# of Topics	Twitter	Reddit
5	0.3372275406	0.4339097835
6	0.3303525918	0.5504873718
7	0.3663472473	0.5211998186
8	0.2914792836	0.4803693326
9	0.3209455841	0.4886859048
10	0.2981905724	0.4453635286
11	0.320238440	0.4792650113
12	0.3262399450	0.4676237023
13	0.3143348988	0.4715463116
14	0.3021454240	0.4891566846
15	0.3243869562	0.4286690553
16	0.308381862	0.4710893458
17	0.2935348829	0.5355977425
18	0.2964168261	0.4160669072
19	0.3070941892	0.5050704624
20	0.3052084118	0.4553972720

Degree centrality, for instance, refers to the actor with the most connections to other actors. The higher the betweenness centrality, the better the actor's position in the network and the easier it is to access resources in the network without relying on other actors. Closeness centrality, on the other hand, measures the proximity of an actor to other actors and represents the distance between an actor and all other actors in the entire network, taking into account the tangentially connected relationships within the network. While degree centrality emphasizes direct connections, closeness centrality emphasizes the length of paths connecting actors. Betweenness centrality acts as an intermediary to control relationships between actors that are not directly connected by computing over the entire network.

Finally, eigenvector centrality is an extension of connection centrality that focuses on the degree of connectedness between connected actors, indicating how important the connected actors are. In other words, actors can be given a weighted centrality to determine which actors have the strongest influence. Overall, the application of network analysis, along with centrality analysis, can provide a more detailed understanding of the structure and relationships within a network, enabling researchers to identify key actors and their influence within the network.

This study utilized the Textnet library to produce a graph modeling language implemented in Python, which yielded a GML file that portrayed the relationships between the trained words. Notably, to minimize the complexity of the analysis process, only the most frequently used 100 words were trained. Subsequently, we employed Gephi, a graphical user interface tool designed for graph analysis, to examine various metrics such as degree centrality. To concentrate on the crucial actors and their influence, we limited our focus to the top 20 words for each centrality analysis criterion.

3. Results

3.1. Word Frequency Analysis

Tables 6 and 7 present the top 100 list generated from the frequency analysis. To avoid repetition, we synonymized words that had the same meaning in context. For instance, “crypto” is often used to refer to “cryptocurrency”, and “giveaway” is frequently used to denote the distribution of tokens or NFTs for free, which we synonymized with “airdrop”.

Table 6. Top 100 frequency words from Twitter.

No.	Freq. Word	No.	Freq. Word	No.	Freq. Word	No.	Freq. Word	No.	Freq. Word
1	nft	21	xhashtag	41	lets	61	participate	81	trading
2	airdrop	22	decentralized	42	floki	62	price	82	ada
3	cryptocurrency	23	luna	43	bullish	63	companionto	83	listing
4	defi	24	event	44	mint	64	elonmusk	84	shib
5	ethereum	25	cmpn	45	earn	65	support	85	polygon
6	community	26	staking	46	sale	66	holders	86	tipmeacoffee
7	web3	27	winner	47	buy	67	avax	87	read
8	blockchain	28	glodao	48	wallet	68	money	88	doge
9	bnb	29	ido	49	campaign	69	presale	89	invite
10	metaverse	30	utc	50	ecosystem	70	gaming	90	vidt
11	token	31	top	51	fortprotocol	71	market	91	technology
12	join	32	eventdao	52	p2e	72	vote	92	social
13	bitcoin	33	cisla	53	altcoin	73	ama	93	claim
14	solana	34	tag	54	swap	74	socialfi	94	public
15	launch	35	game	55	exchange	75	glodaoofficial	95	profit
16	hypernation8	36	chance	56	usdt	76	meme	96	opensea
17	reward	37	governance	57	utility	77	finance	97	daoverse
18	binance	38	people	58	dapp	78	protocol	98	meta
19	gamefi	39	random	59	development	79	coinmarketcap	99	thedaomaker
20	whitelist	40	discord	60	coin	80	yoleeuniverse	100	czbinance

Table 7. Top 100 frequency words from Reddit.

No.	Freq. Word	No.	Freq. Word	No.	Freq. Word	No.	Freq. Word	No.	Freq. Word
1	nft	21	defi	41	locked	61	asset	81	tournament
2	game	22	governance	42	access	62	story	82	metaverse
3	token	23	price	43	mint	63	wallet	83	momentum
4	project	24	blockchain	44	staking	64	vote	84	listed
5	airdrop	25	july	45	market	65	twitter	85	charge
6	community	26	decentralized	46	official	66	stablecoin	86	promotion
7	launch	27	chain	47	coin	67	space	87	rage
8	cryptocurrency	28	reward	48	link	68	latest	88	medical
9	bnb	29	victory	49	world	69	ownership	89	member
10	ethereum	30	platform	50	system	70	lost	90	fueled
11	solana	31	telegram	51	utility	71	cards	91	captain
12	usdq	32	liquidity	52	live	72	connection	92	snippet
13	astrosphere	33	buy	53	tech	73	network	93	rush
14	field	34	event	54	makerdao	74	ape	94	chance
15	apollo	35	usa	55	development	75	dex	95	astrosphenft
16	discord	36	gas	56	protocol	76	majority	96	earn
17	smartcontract	37	ecosystem	57	exclusive	77	dai	97	presale
18	join	38	bitcoin	58	voting	78	tackle	98	supply
19	website	39	value	59	server	79	introduce	99	exchange
20	holder	40	kaiba	60	investment	80	host	100	marketing

NFTs took the top spot in our analysis from both Twitter and Reddit data because we had the most data from 2022, a year in which NFTs were the most popular topic in the blockchain industry. The term “airdrop” also emerged as a prominent keyword on both Twitter and Reddit, likely due to its significance in determining user eligibility to participate in a DAO. To promote their DAOs and attract enthusiastic participants, DAO operators utilize airdrops as a means to grant free access to their platforms.

These airdrops may include NFTs and tokens that hold utility within the DAO’s activities. The allocation of an airdrop, including the selection of recipients and the proportion of total token issuance dedicated to an airdrop, can provide insights into the nature, op-

erational methods, and level of decentralization of the DAO. Major blockchain networks such as Ethereum and Solana were also featured in the top results in Twitter data. On the other hand, in Reddit data, ‘game’ came in second, reflecting the popularity of blockchain gaming, which has been a big topic of interest with P2E (play to earn) alongside NFTs between 2021 and 2022. The terms “lets”, “join”, “participate”, “whitelist”, “reward”, and “earn” highlight the importance of community participation in the establishment of DAOs. Additionally, the words “defi”, “p2e”, “gamefi”, “earn”, “money”, “finance”, and “profit” indicate that DAOs are receiving significant attention in the gaming, finance, and investment sectors.

3.2. Topic Detection and Analysis Using LDA

Table 8 presents the 12 topics identified from Twitter along with the associated terms. The persistent occurrence of the term “community” across different topics suggests that DAOs are inherently designed to enable collaboration and decision-making via blockchain technology. Topic 1 reflects a strong interest in leveraging DAOs in various areas of the blockchain industry, such as NFTs, DeFi, Metaverse, and Games. Many projects are exploring the potential for implementing DAOs in these domains.

Table 8. The result of LDA topic modeling from Twitter.

Topics	Representative Terms
DAO Project (T1)	NFT, Defi, Metaverse, Project, Community, Gamefi, Utility, Team
Social Network (T2)	Glodao, Project, Discord, Team, People, Community
DAO Finance (T3)	Daoverse, Altcoin, Dapp, Staking, Token, Coin, Cryptotrading
DeFi (T4)	DeFi, Bullish, USDT, Earn, Luna, Utility, Vote, Farming
IDO (Initial Dex Offering) (T5)	IDO, DeFi, Entries, Token, Swap, Social, Listing, Dex
DAO Marketing (T6)	Airdrop, Whitelist, Follow, Tag, Friend, Winner, Launch
NFT Project (T7)	Platform, Presale, Utility, Partnership, Governance
Hypernation (T8)	Hypernation, Reward, Twitter, Campaign, Genesis, Winner
Startup (T9)	Cryptocurrency, Startup, Project, Community, Domain, Industry
DAO Management (T10)	Decentralized, Space, Event, World, Citizen, Freedom
Crowdfunding (T11)	Top, Companion, Protocol, Supply, Gleam, Burn
DAO Community (T12)	Member, Launch, Announce, Community, Finance, Project

Topics 2 and 12 highlight how DAOs function as social networks or how they relate to social networks. As DAOs bring together people with common interests or goals to facilitate discussions and decision-making, they have the potential to evolve into communities with chat features and more in the future. Topics 3, 4, 5, and 11 are related to DAO’s financial roles, such as DeFi and IDO, and crowdfunding, which was the original use case for the first DAO, still has significant applications.

In contrast to Twitter, where discussions predominantly centered on DAO projects and operations, marketing, or fundraising, gaming topics emerged as the dominant theme on Reddit. In Table 9, of the eight topics identified, three were directly related to games and ‘game’ was a recurring word in almost all of them.

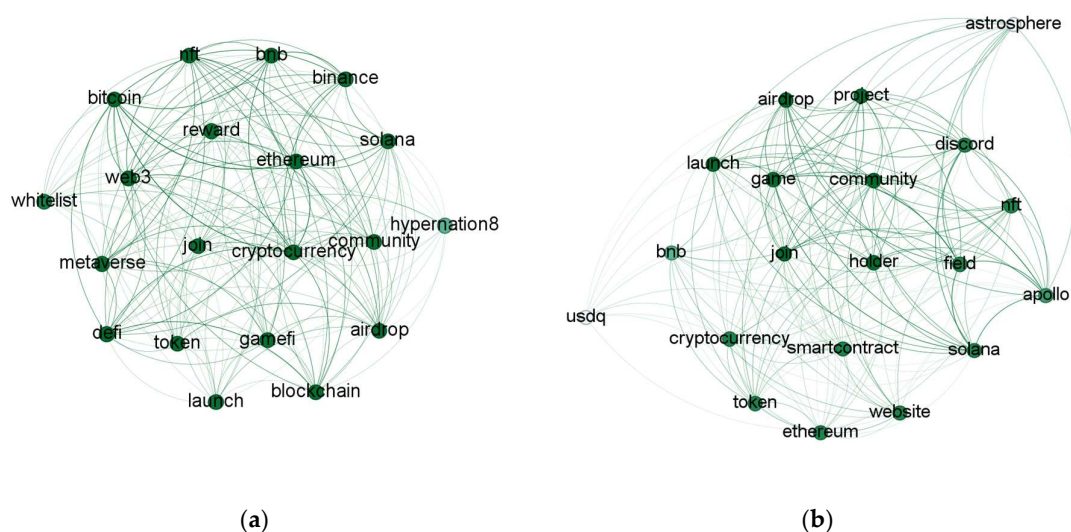
Table 9. The results of LDA topic modeling from Reddit.

Topics	Representative Terms
NFT Project (T1)	Token, NFT, Project, Launch, Community, Cryptocurrency, Mint
DAO Ecosystem (T2)	Cryptocurrency, Decentralized, Governance, Price, Airdrop, Ecosystem
Launchpad (T3)	Game, Project, Launch, Community, Airdrop, Rush, Exclusive
Game Project (T4)	Airdrop, NFT, Launch, Game, Explorer, Tournament, Captain
Stablecoin (T5)	Makerdao, DAI, Token, USDQ, Value, Wallet, Stablecoin
Game Community (T6)	Game, Discord, Field, Victory, Spoiler, Story, Server
Governance (T7)	Trading, GT (Governance Token), Cryptocurrency, Token, Price, Market
Crypto Funding (T8)	Apollo, Game, Point, Token, Project, Holder, Promotion

For example, launchpads, also known as crypto incubators, are decentralized exchange-based platforms that provide funding for crypto projects before they are publicly listed. They are particularly popular for gaming project launches. Unlike Twitter, Reddit has seen the rise of stablecoins such as DAI and USDQ, which can be seen as a way to incentivize participants in cryptocurrency-based games based on their ability to maintain a stable value. Topics such as DAO ecosystems and governance have also emerged, potentially reflecting platform-specific characteristics of Twitter and Reddit. Twitter, as an open social media platform, is commonly used for marketing, promotion, and announcements, whereas Reddit is a closed community where individuals engage in specific topic discussions and debates.

3.3. Network Analysis

Figure 7 presents a visual representation of the interrelationships of the top 20 keywords within the set of the 100 most frequently occurring keywords. The figure provides a graphical representation that offers insights into the connections and associations between these key terms, allowing for a better understanding of their interplay and significance within the context of the study. Furthermore, we conducted a comprehensive analysis by calculating several centrality measures, namely degree centrality, eigenvector centrality, betweenness centrality, and closeness centrality, for the top 100 keywords.

**Figure 7.** (a) The result of network analysis from Twitter; (b) the result of network analysis from Reddit.

Centrality measures play a crucial role in analyzing networks. For example, degree centrality assesses the importance of a node based on the number of links it possesses. It helps identify highly connected individuals, popular figures, or those likely to possess a significant amount of information, enabling them to quickly connect with the broader network. Similarly, eigenvector centrality evaluates a node's influence by considering both its number of connections and the quality of those connections.

It considers the well-connectedness of the node's connections and extends this analysis throughout the network. On the other hand, betweenness centrality quantifies the frequency with which a node lies on the shortest paths between other nodes. Lastly, closeness centrality assigns scores to nodes based on their proximity to all other nodes in the network. By employing closeness centrality, we can identify individuals who hold strategic positions to exert influence over the entire network more efficiently.

The resulting centrality scores allowed us to identify the keywords that occupied prominent positions within the network. Specifically, Tables 10 and 11 present the top 20 keywords with the highest centrality scores, showcasing their significance in the context of the study.

In the case of Twitter, the results of the frequency analysis revealed that several words, such as nft, airdrop, cryptocurrency, defi, ethereum, community, web3, blockchain, and bnb, ranked high in the network analysis, as well. This indicates that these words are central to the network, based on various criteria. Interestingly, czbinance, which was positioned as the 100th most frequent word in the analysis, emerged as the 6th most central word based on eigenvector centrality, betweenness centrality, and closeness centrality measures. Similarly, elonmusk, despite being ranked 64th in frequency analysis, placed within the top 20 in all centrality measures except degree centrality. These findings illustrate the significant influence of individuals such as Binance founder Changpeng Bellavitis and Elon Musk, even though they are not relatively frequently mentioned.

In contrast to Twitter, the analysis of Reddit (Table 11) revealed that the top-ranking words in degree centrality displayed divergent results in the other three centrality measures. This discrepancy may stem from the nature of Reddit as a more self-contained platform, where users are able to compose lengthier posts compared to Twitter. Consequently, certain words with relatively lower impact or significance may garner higher frequency and connectivity within the network.

Table 10. Centrality analysis from Twitter.

Rank	Degree Centrality	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality
1	cryptocurrency	bnb	bnb	bnb
2	nft	cryptocurrency	cryptocurrency	cryptocurrency
3	bitcoin	nft	nft	nft
4	ethereum	bitcoin	bitcoin	bitcoin
5	bnb	web3	web3	web3
6	web3	czbinance	czbinance	czbinance
7	defi	ethereum	ethereum	ethereum
8	solana	defi	defi	defi
9	binance	airdrop	tag	tag
10	luna	blockchain	elonmusk	elonmusk
11	airdrop	community	metaverse	metaverse
12	metaverse	token	solana	airdrop
13	staking	event	binance	blockchain
14	companionto	winner	coinmarketcap	community
15	blockchain	support	exchange	token
16	exchange	lets	wallet	event
17	swap	read	buy	winner
18	wallet	earn	airdrop	support
19	community	metaverse	blockchain	lets
20	bullish	elonmusk	community	read

Table 11. Centrality analysis from Reddit.

Rank	Degree Centrality	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality
1	field	community	community	community
2	astrosphere	launch	launch	launch
3	solana	project	project	project
4	victory	airdrop	airdrop	airdrop
5	apollo	game	links	link
6	game	holder	nft	nft
7	project	event	game	game
8	july	access	holder	holder
9	discord	majority	event	event
10	airdrop	latest	access	access
11	holder	lost	majority	majority
12	community	story	latest	latest
13	launch	join	lost	lost
14	link	official	story	story
15	tackle	nft	join	join
16	card	member	official	official
17	official	link	token	token
18	event	token	smartcontract	smartcontract
19	access	smartcontract	voting	voting
20	exclusive	voting	buy	buy

4. Discussion

This study has produced several noteworthy findings. Firstly, the analysis of the top 100 keywords identified prominent mainnet coins, namely bitcoin, ethereum, bnb, solana, avax, and polygon. While existing research on DAOs and blockchain technology often emphasizes the security and scalability aspects of blockchains [26–29], the selection of a mainnet for establishing and operating a DAO is crucial. Each mainnet possesses distinct features that can influence the characteristics of an organization, warranting further investigation to determine the most suitable mainnet for specific DAOs. For instance, different mainnets may impose varying time and financial requirements for proposals and voting processes, which can serve as vital variables impacting the functioning of a DAO and the involvement of its members.

The second significant finding pertains to the identification of key themes that reflect the sectors and types of organizations displaying the most interest in DAOs. This study has revealed a notable level of enthusiasm and utilization of DAOs in domains such as NFTs, gaming, and finance. While previous discussions on the application of blockchain technology in organizations have primarily revolved around government and corporate governance studies [30–33], the present research highlights the growing adoption of DAOs within specific sector-specific projects, particularly in gaming and finance.

This observation aligns with the statistics provided by Deepdao, which provides the real-time statistics on DAOs worldwide, where the leading treasury projects predominantly consist of financial organizations, with the exception of layers 1 and 2. Such a phenomenon can be attributed to one of the distinguishing characteristics of DAOs, which grants ownership to all project members and allows direct participation in crucial decision-making processes, including operational procedures and compensation policies.

Consequently, it is only natural for projects and organizations operating in sectors such as gaming and finance, where organizational policies directly impact the economic interests of their members, to exhibit a strong interest in embracing DAOs. The study also revealed that the most common words in the blockchain industry, such as nft, airdrop, cryptocurrency, defi, ethereum, community, web3, blockchain, and bnb, are highly central and influential. However, it is interesting to note that individuals such as ‘elonmusk’ and ‘czbinance’ wield significant influence, despite their relatively low frequency of occurrence

in the text. The prominence of influential figures in discussions and narratives surrounding DAOs can be attributed to the nascent stage and relatively small size of the DAO market.

During these early stages of market formation, the stories and actions of key individuals have a higher likelihood of being widely discussed and circulated, and of ultimately exerting influence on the market. As the DAO ecosystem continues to evolve and mature, it is anticipated that a broader range of factors and dynamics will shape the market landscape, thereby diversifying the narratives and sources of influence within the industry.

5. Conclusions

This study is expected to stimulate further research and discussions on the utility and challenges of DAOs across various academic, industrial, and institutional domains. Based on the study's findings, a few research directions emerge for further exploration. First, investigating the evolving landscape of DAOs and their impact on various industries and sectors can provide valuable insights into their long-term potential and challenges. Understanding how DAOs can reshape governance, participation, and economic models across different domains is crucial for informed decision-making and strategic planning. Second, examining the scalability and security aspects of blockchain technology, specifically within the context of DAOs, is essential. As DAOs continue to gain prominence, ensuring the scalability, efficiency, and robustness of underlying blockchain infrastructures becomes paramount.

Research focusing on addressing technical barriers and exploring innovative solutions to enhance the scalability and security of DAOs can pave the way for their widespread adoption. Additionally, it is essential to address the legal and ethical issues associated with DAOs. The decentralized nature of DAOs raises questions about regulatory frameworks, accountability, and dispute resolution mechanisms.

Future research should explore the legal implications of DAO operations, including governance structures, member responsibilities, and compliance with existing laws and regulations. Additionally, ethical considerations regarding DAO decision-making, transparency, and inclusivity should be examined to ensure the responsible and sustainable development of DAOs.

Finally, exploring the social and economic implications of DAOs is crucial. This entails investigating the impact of DAOs on traditional organizational structures, employment patterns, and economic systems. Understanding the potential benefits, challenges, and unintended consequences of DAOs can guide policymakers, industry leaders, and stakeholders in harnessing the full potential of this transformative technology.

It is important to acknowledge the limitations of this study. One significant limitation is the lack of temporal consideration. Given that the majority of data collection from Twitter and Reddit occurred in 2022, it is plausible that the prominence of specific keywords and topics during that period may have exerted an influence on the overall findings. Consequently, it is imperative that future research endeavors encompass a temporal dimension to gain a more comprehensive understanding of the dynamic nature of DAOs and their associated subjects.

Additionally, the methodology employed in analyzing the data, including centrality metrics and keyword frequency analysis, provides a quantitative perspective on the topic. While these approaches offer valuable insights, they may overlook nuanced qualitative aspects of DAOs, such as individual experiences, motivations, and cultural factors that shape the adoption and utilization of DAOs. Incorporating qualitative research methods, such as interviews and case studies, could provide a more comprehensive understanding of the complexities surrounding DAOs.

Furthermore, the study's focus on the English language limits the generalizability of the findings to a global context. DAOs are a global phenomenon, and their adoption and discussions extend beyond English-speaking communities. Including data from multiple languages and cultural contexts would enhance the study's breadth and capture a more diverse range of perspectives and trends.

The unique characteristics of these two platforms resulted in variations in the results. However, this study did not delve deeply into analyzing these differences on an individual basis. Therefore, future qualitative research, such as case studies, would be valuable in complementing the findings of this study and providing a more comprehensive understanding of the nuanced variations observed between Twitter and Reddit in relation to DAOs.

Lastly, this study was a basic trend analysis study based on social data, which limited the in-depth discussion of data analysis techniques. Therefore, in future research, it is necessary to develop algorithms or conduct more advanced research considering various data analysis techniques based on DAO-related company big data collection.

Author Contributions: Conceptualization, H.P.; methodology, H.P.; software, H.P.; validation, B.K.; formal analysis, H.P. and B.K.; investigation, H.P. and B.K.; resources, H.P.; data curation, H.P.; writing—original draft preparation, H.P. and B.K.; writing—review and editing, B.K. and I.U.; visualization, B.K. and I.U.; supervision, B.K. and I.U.; project administration, B.K.; funding acquisition, H.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable for data availability.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Aggarwal, S.; Kumar, N. Blockchain 2.0: Smart contracts. *Adv. Comput.* **2021**, *121*, 301–322.
- Deepdao. Available online: <https://deepdao.io/organizations> (accessed on 7 May 2023).
- Dhillon, V.; Metcalf, D.; Hooper, M.; Dhillon, V.; Metcalf, D.; Hooper, M. The DAO hacked. *Blockchain Enabled Appl.* **2017**, *1*, 67–78.
- Dasgupta, D.; Shrein, J.M.; Gupta, K.D. A survey of blockchain from security perspective. *J. Bank. Financ. Technol.* **1991**, *3*, 1–17. [\[CrossRef\]](#)
- Mehar, M.I.; Shier, C.L.; Giambattista, A.; Gong, E.; Fletcher, G.; Sanayhie, R.; Kim, H.M.; Laskowski, M. Understanding a revolutionary and flawed grand experiment in blockchain: The DAO attack. *J. Cases Inf. Technol.* **2019**, *21*, 19–32. [\[CrossRef\]](#)
- Singh, M.; Kim, S. Blockchain technology for decentralized autonomous organizations. *Adv. Comput.* **2019**, *115*, 115–140.
- Faqir, Y.; Arroyo, J.; Hassan, S. An overview of decentralized autonomous organizations on the blockchain. In Proceedings of the 16th International Symposium on Open Collaboration, Virtual, 25–27 August 2020; pp. 1–8.
- Liu, L.; Zhou, S.; Huang, H.; Zheng, Z. From technology to society: An overview of blockchain-based DAO. *IEEE Open J. Comput. Soc.* **2021**, *2*, 204–215. [\[CrossRef\]](#)
- Beck, R.; Müller-Bloch, C.; King, J.L. Governance in the blockchain economy: A framework and research agenda. *J. Assoc. Inf. Syst.* **2018**, *19*, 1–8. [\[CrossRef\]](#)
- Diallo, N.; Shi, W.; Xu, L.; Gao, Z.; Chen, L.; Lu, Y.; Shah, N.; Carranco, L.; Le, T.-C.; Surez, A.B. eGov-DAO: A better government using blockchain based decentralized autonomous organization. In Proceedings of the 2018 International Conference on eDemocracy & eGovernment (ICEDEG), Ambato, Ecuador, 4–6 April 2018; pp. 166–171.
- Santos, F.; Kostakis, V. *The DAO: A Million Dollar Lesson in Blockchain Governance*; School of Business and Governance, Ragnar Nurkse Department of Innovation and Governance: Tallinn, Estonia, 2018.
- Morrison, R.; Mazey, N.C.; Wingreen, S.C. The DAO controversy: The case for a new species of corporate governance? *Front. Blockchain* **2020**, *3*, 25–48. [\[CrossRef\]](#)
- Ding, W.W.; Liang, X.; Hou, J.; Wang, G.; Yuan, Y.; Li, J.; Wang, F.-Y. Parallel governance for decentralized autonomous organizations enabled by blockchain and smart contracts. In Proceedings of the 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPi), Beijing, China, 15 July–15 August 2021; pp. 1–4.
- Kaal, W.A. Decentralized corporate governance via blockchain technology. *Ann. Corp. Gov.* **2020**, *5*, 101–147. [\[CrossRef\]](#)
- Saurabh, K.; Rani, N.; Upadhyay, P. Towards blockchain led decentralized autonomous organization (DAO) business model innovations. *Benchmarking: Int. J.* **2023**, *30*, 475–502. [\[CrossRef\]](#)
- Chughtia, Z.A.; Awais, M.; Rasheed, A. Distributed autonomous organization security in blockchain:(DAO attack). *Int. J. Comput. Innov. Sci.* **2022**, *1*, 47–59.
- Abraham, J.; Higdon, D.; Nelson, J.; Ibarra, J. Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Sci. Rev.* **2018**, *1*, 1–21.
- Bala, D.E.; Stancu, S. Using Twitter Data and Lexicon-Based Sentiment Analysis to Study the Attitude towards Cryptocurrency Market and Blockchain Technology. In Proceedings of the Education, Research and Business Technologies: Proceedings of 21st International Conference on Informatics in Economy (IE 2022), Singapore, 1 January 2023; pp. 187–198.
- Gao, W.; Sebastiani, F. From classification to quantification in tweet sentiment analysis. *Soc. Netw. Anal. Min.* **2016**, *6*, 19. [\[CrossRef\]](#)

20. Rouhani, S.; Abedin, E. Crypto-currencies narrated on tweets: A sentiment analysis approach. *Int. J. Ethics Syst.* **2020**, *36*, 58–72. [[CrossRef](#)]
21. Fox, N.; Graham, L.J.; Eigenbrod, F.; Bullock, J.M.; Parks, K.E. Reddit: A novel data source for cultural ecosystem service studies. *Ecosyst. Serv.* **2021**, *50*, 101331. [[CrossRef](#)]
22. Yeajin, J.; Yoonjae, N. Trends and important studies in Korean tourism research: A social network analysis of citations in a leading journal (2014–2016). *Int. J. Tour. Manag. Sci.* **2018**, *33*, 1–42.
23. Newman, D.; Bonilla, E.V.; Buntine, W. Improving topic coherence with regularized topic models. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 1–9.
24. Lee, S. A content analysis of journal articles using the language network analysis methods. *Korea Soc. Inf. Manag.* **2014**, *31*, 49–68.
25. Das, K.; Samanta, S.; Pal, M. Study on centrality measures in social networks: A survey. *Soc. Netw. Anal. Min.* **2018**, *8*, 13. [[CrossRef](#)]
26. Jun, M. Blockchain government—A next form of infrastructure for the twenty-first century. *J. Open Innov. Technol. Mark. Complex.* **2018**, *4*, 7. [[CrossRef](#)]
27. Wang, S.; Ding, W.; Li, J.; Yuan, Y.; Ouyang, L.; Wang, F.Y. Decentralized autonomous organizations: Concept, model, and applications. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 870–878. [[CrossRef](#)]
28. Zhao, X.; Ai, P.; Lai, F.; Luo, X.; Benitez, J. Task management in decentralized autonomous organization. *J. Oper. Manag.* **2022**, *68*, 649–674. [[CrossRef](#)]
29. Hsieh, Y.Y.; Vergne, J.P.; Anderson, P.; Lakhani, K.; Reitzig, M. Bitcoin and the rise of decentralized autonomous organizations. *J. Organ. Des.* **2018**, *7*, 14. [[CrossRef](#)]
30. Bellavitis, C.; Fisch, C.; Momtaz, P.P. The rise of decentralized autonomous organizations: A first empirical glimpse. *Ventur. Cap.* **2023**, *25*, 187–203. [[CrossRef](#)]
31. Murray, C.; Albareda, L. Blockchain and the emergence of Decentralized Autonomous Organizations (DAOs): An integrative model and research agenda. *Technol. Forecast. Soc. Change* **2022**, *182*, 121806.
32. Murray, A.; Kuban, S.; Josefy, M.; Anderson, J. Contracting in the smart era: The implications of blockchain and decentralized autonomous organizations for contracting and corporate governance. *Acad. Manag. Perspect.* **2021**, *35*, 622–641. [[CrossRef](#)]
33. Norta, A. Designing a smart-contract application layer for transacting decentralized autonomous organizations. In Proceedings of the Advances in Computing and Data Sciences: First International Conference, ICACDS 2016, Ghaziabad, India, 11–12 November 2017; pp. 595–604.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.