Sustainability Assessment of Ethereum Developers Issues and Comments: Topic and Network Analysis

M. Vaccargiu[1,3], S. Aufiero[2], S. Bartolucci[2], R. Neykova[3], R. Tonelli[1], G. Destefanis[3]
[1]University of Cagliari, Italy,
[2]University College London, UK,
[3]Brunel University London, UK
{matteo.vaccargiu,roberto.tonelli}@unica.it,
{sabrina.aufiero.22, s.bartolucci}@ucl.ac.uk,
{rumyana.neykova,giuseppe.destefanis}@brunel.ac.uk

***Abstract*—Blockchain technology presents several sustainability challenges, such as optimization problems and excessive energy usage, which developers may overlook. This study examines how blockchain developers incorporate sustainability themes into their conversations and the specific areas they focus on. We analyze `Go-Ethereum` developers' issues and comments from Github to identify key sustainability topics using the BERT model. Our methodology examines themes related to energy use and sustainability, considering five dimensions: economic, social, individual, environmental, and technical, based on the Sustainability Awareness Framework (SusAF). We also analyze the network of developers engaged in these topics using centrality measures and community detection. This reveals key influencers, developer clusters, and trends in sustainability discussions. Our findings show a dominance of technical discussions, with a growing focus on economic and environmental sustainability. This analysis highlights key sustainability areas, identifies gaps, and suggests factors to ensure the development of more sustainable blockchain technologies.**

***Index Terms*—Blockchain, Topic Modelling, NLP, BERT, SusAF, Sustainability, Energy, Gas Consumption, Optimization, Network Analysis**

I. Introduction

Since its introduction, blockchain technology has witnessed rapid growth and widespread adoption, attributed to its key features such as security, scalability, transparency, immutability, and traceability [1]. These features have driven the adoption of blockchain technologies in a wide range of industries.

However, the surge in popularity of cryptocurrencies, especially Bitcoin, along with other blockchain implementations, has raised significant concerns regarding their environmental footprint. This is primarily due to the increased greenhouse gas emissions and substantial energy consumption associated with their operations [2]. Such environmental concerns have sparked ongoing debates within both scientific communities and industry practices about the overall sustainability of blockchain and distributed ledger technologies [3]. A particular point of contention is the extensive energy requirements driven by the computational processes they employ, notably the Proof-of-Work (PoW) algorithm. This algorithm, commonly used in cryptocurrency mining, demands considerable computing resources [4], highlighting the urgent need for a reassessment of the energy efficiency of these technologies [5].

This study aims to understand how developers prioritize and address sustainability concerns in their projects. Analyzing developer discussions is crucial because developers shape the evolution of blockchain technology. These discussions highlight practical challenges in implementing sustainability measures and can indicate future trends toward sustainable practices. They also reveal the tension between technical optimization and environmental concerns, providing context for policymakers and researchers. Understanding these conversations can inform strategies to promote eco-friendly practices in blockchain development and influence the future environmental impact of blockchain technology.

Topic modelling offers a powerful tool to sift through textual data to discover the main themes of a discussion. This research explores a dataset comprising issues and comments from Go-Ethereum developers and the network of *sustainability-related* developers. The Go-Ethereum command-line interface enables developers to operate full Ethereum blockchain nodes and deploy smart contracts. Our aim is to introduce a reproducible approach for analysing the content of discussions among developers, assessing the sustainability of the topics they engage with, and tracking any developmental trends over time. This study is guided by the following research questions:

RQ1: Are sustainability issues a topic of discussion among Ethereum developers?
Rationale: Understanding whether sustainability is discussed helps assess the community's awareness of and engagement with environmental and ethical issues. This is important for evaluating potential future sustainability improvements in blockchain technologies. Our findings indicate that while developers have started addressing these issues, their efforts are not yet adequate.

RQ2: How do sustainability-related discussions evolve over time?
Rationale: Analyzing the evolution of sustainability discussions can reveal changing priorities within the developer community and their responsiveness to global trends. This provides insights into shifts in developer

attitudes and helps anticipate future project directions, important for understanding the trajectory of blockchain sustainability efforts. The analysis shows that the proportion of sustainability-related issues peaked in recent years, especially around the *Ethereum Merge* event in September 2022.

RQ3: Which sustainability-related developers are the most influential? Are they also among the most influential across the entire Go-Ethereum network?
Rationale: Identifying key influencers in sustainability discussions highlights individuals who shape these conversations and influence sustainable blockchain development. Understanding their impact in both sustainability discussions and the broader network shows how sustainability concerns are integrated into the development process. These users contribute to discussions and facilitate collaboration and information exchange.

RQ4: Do developers engage in any active topic, or do they select topics based on their specific domains (e.g., social vs. environmental discussions)?
Rationale: Exploring whether developers specialize in certain topics helps understand how sustainability concerns are distributed across the community. This is essential for identifying gaps in expertise and ensuring comprehensive coverage of blockchain sustainability. Our findings indicate that developers prefer discussions similar to their previous engagements, rather than responding randomly to any open issue.

This study extends our previous work titled *Sustainability in Blockchain Development: A BERT-Based Analysis of Ethereum Developers' Discussions* [6] presented at EASE 2024[1]. In this extended version, we have expanded our research to include RQ3 and RQ4. These additional questions allow us to further investigate the identification of influential developers in sustainability discussions and to examine whether developers engage with specific topics based on their domain focus.

The paper is organised as follows: in Section II we present the existing literature connected to this study; then, in Section III we discuss in detail the methodology. The results are presented and discussed in Section IV, along with the answer to the research questions. Threats to validity are presented in Section V. Finally, in Section VI, conclusions and future research developments in this area are discussed.

The code used for this study is available at this *Code Link*.

II. Related works

The analysis of sustainability and energy consumption of blockchain technology has attracted significant attention from the scientific community. This section outlines the key contributions from recent literature, comparing them with some of the of the objectives and methodologies of this study.

Qin and Gervais in [7], Arshad et al. in [8], and Asif et al. in [9] provide general analysis related to energy consumption and sustainability in Ethereum. The authors discuss platform's consumption, possible solutions and challenges to improve it, as well as the transition from proof-of-work to proof-of-stake protocol.

An investigation about agile blockchain-oriented software development principles and sustainability software design principles was conducted by Pinna et al. in [10], where they present a new Agile method for the development of blockchain-oriented systems that includes sustainability awareness practices within the development phases, in particular in the requirements and the acceptance tests.

Eligüzel in [11] presents a study on the relationship between blockchain technology and sustainability through a descriptive literature review, using topic modelling and clustering method of latent semantic analysis (a social spider optimization technique) on the corpus of 1069 articles extracted from Scopus.

Another literature data mining work was carried out by Liu et al. in [12], where 759 articles extracted from Web of Science related to blockchain technology in sustainable financial field were analysed by keyword analysis, bi-clustering algorithms, and strategic coordinate analysis so as to explore the hot topics in this field and predict the trend of future sustainable development. Also, related to the sustainability of blockchain with applications in finance, a comparison using a holistic approach between the old and new ways financial transactions was conducted by Stamoulis in [13], analysing their sustainability performance.

Ayman et al. in [14] created a topic analysis model considering data extracted from the smart contract developer community on Stack Overflow, providing information on the topics and issues discussed between the users.

Other studies on blockchain topic analysis are more focused on using BERT-based models to analyse generic blockchain-related topics with data extracted from social media such as Reddit [15], or based on abstracts gathered from USPTO patents [16], highlighting in both cases the benefits of using an NLP-based BERT textual analysis approach to examine technological knowledge and relationships within the field of blockchain technology. Complex networks approaches to analyse blockchain systems have been widely used in the context of understanding crypto markets [17], [18], investment networks [19], transactions patterns [20], and smart contracts dependencies [21]–[26] to mention a few. Regarding the analysis of the network of Ethereum developers, to the best of our knowledge there are no specific works in the literature that consider the interaction between them based on issues and comments extracted from GitHub. Previous literature has focused on the interplay between developers' sentiment on Github and prediction of returns in crypto markets [27]–[29]. Other approaches similar to ours focus more generally on the analysis of Open-Source Software (OSS). Some studies of GitHub issues and comments related to OSS are proposed by Mumtaz et al. in [30], with the feature "assign issues to issue commenters", and by Jamieson et al. in [31], which considered also commits related to Decentralized Web communities. They analyse respectively 13 and 52 projects, aiming to exam the social smells in the software teams before and after the introduction of this new feature.

In this context, our work proposes a novel approach by applying topic and developers' network analysis of a particu-

[1] https://conf.researchr.org/home/ease-2024

lar OSS example such as Go-Ethereum, providing important insights for the analysis of similar ecosystems.

III. Methodology

This study proposes a reliable and reproducible NLP-based approach to analyse discussion topics. The focus is on identifying and providing insights into the most frequently addressed themes by Ethereum developers in GitHub issues and comments, especially those pertaining to sustainability. In this section, we provide details of the dataset used, including its statistical features, and explain the method we employed to extract and analyse the topics.

A. Dataset overview and statistics

The dataset employed in this research was obtained from GitHub, and centers on discussions related to Go-Ethereum spanning from 2014 up to May 2023. Our analysis focuses on the text of issues and comments linked to them, encompassing a total of 15,954 issues and 50,023 comments. For every issue, the dataset captures several pieces of information, including an ID, the name of the author, the count of comments, the date of the first posting, the date of the latest update, and the full text of the issue. In a similar manner, each comment is detailed by the author's ID, the ID of the associated issue, the date it was created, and the text of the comment itself. In Table I we show an example of issue and in Table II we present an example of comment.

We have conducted data pre-processing to ensure data quality as detailed below. We removed `NaN` values and stop words to reduce noise and improve model's performance. A manual sampling of 600 comments revealed that code snippets were primarily found in issue bodies rather than comments. To minimise the impact of non-natural language content on topic modelling, we merged the comments with their corresponding issues into a single column. In this way, we could capture the complete context of the discussions, while reducing the influence of code snippets on the model's performance.

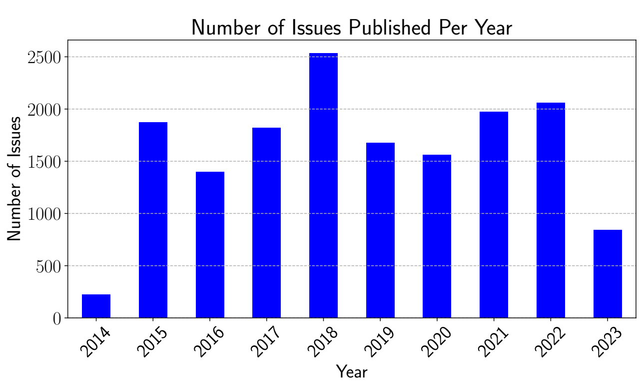


Fig. 1: Number of issues published per year from 2014 to 2023.

Figure 1 shows the year-by-year breakdown of issues posted on GitHub. Omitting 2014, the study's start year, and 2023, for which data is only available until May, the yearly distribution of issues is fairly even. The number of issues annually varied from $1,398$ in 2016 to $2,061$ in 2022, with the highest activity observed in 2018, featuring $2,534$ issues. This justifies the following analysis given by year. Figure 2 shows the distribution of the lifespan, in years, of closed issues. The median time to resolution is 6 days, showing that half of the issues were addressed within a week. Additionally, 75% of the issues were closed in 83 days or fewer. Out of the total $15,954$ issues, only 241 are still *open*. These open issues generally last much longer than the ones that have been closed, with the median open time showing that half of them stay unresolved for at least 125 days. This may indicate that the issues that remain open tend to be either more complex or contentious.

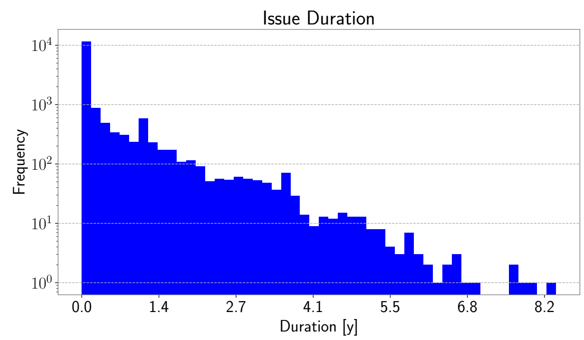


Fig. 2: Lifespan of issues marked as *closed* in the dataset.

To each issue, we assign a measure of *discussion intensity* based on the number of comments associated with it. The distribution of this metric is plotted in Figure 3. It is important to note that while a high number of comments can indicate significant developer interest, it may also reflect the complexity or contentiousness of an issue, or simply a prolonged resolution process. Figure 3 shows that 75% of the issues receive 4 comments or less, indicating that most discussions are brief. The issue with the highest number of comments (470) was titled "Rinkeby Faucet Down," which remained open for 186 days before being closed on March 19, 2022. This high comment count likely reflects both the issue's impact on developers' testing processes and the extended time required for its resolution. This metric provides insight into the level of discussion and potentially the complexity or importance of the issue within the development process.

The data shows that while most issues are resolved quickly with minimal discussion, a few require significant community attention and take longer to address. We analyzed user activity to identify the five most active contributors based on the number of issues they initiated. The top contributors, with user IDs 129561, 142290, 5959481, and 6264126, opened 1208, 863, 669, 455, and 426 issues, respectively. These individuals are also among the top commentators, showing their substantial engagement with the platform. We calculated user lifespan as the time from their first to last recorded activity. On average, users are active for 200 days, but the median lifespan is only 1 day, indicating most users post a single comment and then stop participating. This highlights the brief nature of participation for many users in this ecosystem.

id	number	title	user_fk	repo_fk	state
1720493704	27322	website: Matomo update for cloud migration	54227730	15452919	closed
locked	**comments**	**created_at**	**updated_at**	**author_association**	**body**
false	3	2023-05-22 20:41:03	2023-06-06 19:38:11	MEMBER	”## Description - Updates the Matomo settings to include new secondary tracker for Matomo Cloud migration - ‘.env.local.example‘ updated, along with Netlify env vars”

TABLE I: Issue example

id	user_fk	issue_fk	created_at	author_association	body
1557837189	142290	1720053500	2023-05-22 19:37:58	CONTRIBUTOR	That ”live chat” is a scam. Will ban the account asap. Cc @aldekein

TABLE II: Comment example

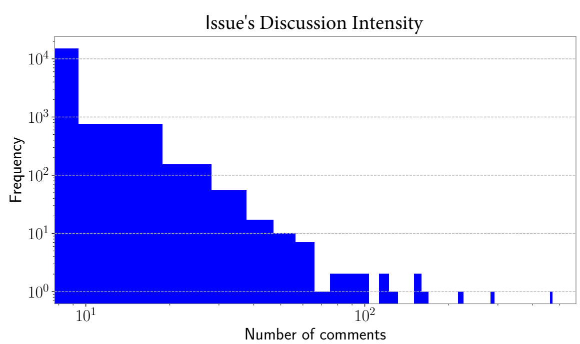


Fig. 3: Discussion intensity: number of comments per issues

[32], [33].

Figure 4 show the steps followed for the dataset generation, topic modelling and topic interpretation. The latter two are explained in detail in the next two subsections.

B. *Topic modelling*

Topic modeling uses natural language processing (NLP) techniques and probabilistic algorithms to identify topics from text. We used the BERT (Bidirectional Encoder Representations from Transformers) model [34] due to its advanced language context handling, efficiency with short texts, and minimal hyperparameter tuning. By employing class-based TF-IDF (c-TF-IDF) [35], BERT effectively associates words with relevant topics, improving clarity and interpretability. This approach produces topics identified by keywords, each with a likelihood score. We used *BERTopic*, which uses BERT's contextual embeddings to analyze topics. The model converts text into uniform hidden representations. Initial trials on the full dataset produced ambiguous results with over 160 varied topics, leading us to adopt a semi-supervised zero-shot method for more precise topic identification.

A semi-supervised zero-shot approach is a hybrid technique that combines the advantages of both supervised and unsupervised learning but does not require explicit examples of every category for training. Instead, it uses known keywords for some data points to infer the classification of unlabelled data, even for categories not seen during training. Employing this technique facilitated a more directed analysis, allowing us to guide the model through the incorporation of issue titles and the specific topics pertinent to our study, thus enhancing the model's ability to classify discussions into relevant and previously undefined topics. We applied *KeyBERT*[2], as discussed by Khan et al. [36], a tool developed specifically for extracting keywords using BERT. This process involves feeding the text from issues and comments into KeyBERT, which then identifies and returns a set of keywords for each document. These keywords are selected based on the phrases within a document that most closely match the document's overall content. The resulting collection of keywords forms a vocabulary that is used as input for CountVectorizer, a component of Sklearn, aiding the BERTopic model in its identification of topics. Furthermore, as described in [37], employing a Zero-Shot classification model with unlabelled data enhances the accuracy of topic extraction.

The model was created using containing the following candidate topics of interest: “consumption”, “consensus”, “cost”, “efficiency”, “energy”, “fee”, “gas”, “green”, “management”, “merge”, “mining”, “optimization”, “pos”, “pow”, “proof-of-stake”, “proof-of-work”, “protocols”, “scalability”, “sustainability”. These were chosen based on the literature reviews provided in Section I and Ethereum's transition from proof-of-work to proof-of-stake, called The Merge [3], which significantly reduced energy consumption and gas emissions [38].

For each experiment, the embedding model used was *BAAI/bge-small-en* [4] which performs well with technical texts and comments as shown by Blanco-Cuaresma et al. in [39]. For the representative model we used *facebook/bart-large-mnli* [5].

To evaluate the effectiveness of our topic model, we used a method that combines topic interpretation with calculating the coherence score c_v as outlined by Roder et al. [40].

[2]https://maartengr.github.io/KeyBERT/api/keybert.html
[3]https://ethereum.org/en/roadmap/merge/
[4]https://huggingface.co/BAAI/bge-small-en
[5]https://huggingface.co/facebook/bart-large-mnli

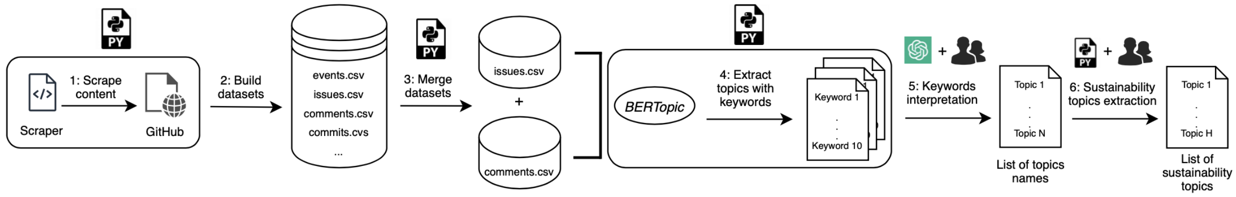


Fig. 4: Research Methodology Applied To Topic Modelling Analysis

The coherence score yielded a value of c_v equal to 0.66, which is generally considered to indicate a satisfactory level of coherence. However, the model generated a total of 165 topics, a figure deemed excessively high. The emergence of numerous small clusters prompted a subsequent experiment with the $min\_topic\_size = 30$ parameter set, aiming to establish a minimum size for each topic cluster.

In this revised model, the c_v coherence value of 0.67 was maintained, but the number of topics decreased to 58, making the results easier to interpret.

A manual review of the topics validated the model's findings, showing them to be consistent.

C. Topic interpretation

In the process of topic interpretation, BERT generates a list of 10 keywords for each topic, accompanied by a probability score for each keyword. This score signifies the relevance of the specific word to the given topic. The interpretation was carried out using a dual approach: initially, Chat-GPT 3.5 analysed the keywords and their respective probabilities to identify the subject of discussion. This automated interpretation was subsequently validated through a manual review by the authors [41], [42].

The use of Chat-GPT 3.5 for topic interpretation was motivated by its ability to understand the context and meaning of the keywords, similar to how a human would interpret them. We tailored a prompt that instructed the tool to generate topic labels based on the provided keywords and their probabilities, ensuring a focused and relevant interpretation process. While keywords alone may not always capture the full scope of a topic, they serve as a strong foundation for both human and machine interpretation.

Colavito et al. [42] demonstrated the effectiveness of GPT-like models for automated labeling tasks of issues without the need for fine-tuning. The study showed substantial agreement between GPT-like models and human annotators, suggesting that these models can be used to reduce the costs associated with manual annotation.

To ensure the accuracy of the Chat-GPT 3.5 interpretations, we also conducted a manual review process. This involved examining the assigned labels and comparing them with the keywords and their probabilities. The manual validation step allowed us to confirm that the labels accurately reflected the underlying topics.

After analysing the full list of topics, we focused on isolating those pertinent to sustainability. This selection was based on the labels assigned to each topic by Chat-GPT and the top 10 words for each topic generated by BERTopic. A topic was flagged as relevant to sustainability if its labels or keywords matched any of the predefined subjects of interest in the zero-shot model. Following the identification of all topics associated with sustainability, we conducted a thorough manual review to validate the findings.

D. Sustainability Network Analysis

For our network analysis, we began with a dataset of 1,468 developers and 2,752 issues, constructing a bipartite network with developers as Layer 1 and issues as Layer 2. Links between layers were established based on developers' comments on issues, resulting in an undirected and unweighted initial network. After removing 473 issues without comments, we focused on the remaining 2,279 issues.

We created a bi-adjacency matrix where rows represented developers and columns represented issues, with binary entries indicating whether a developer commented on an issue. The network density is 0.14%, with an average of 3.8 issues commented on per developer.

Next, we construct the projections of the bipartite network on Layer 1 (developers) and Layer 2 (issues) to characterize developers' relevance and thematic importance of issues.

First, we transformed the bi-adjacency matrix into a 1468 x 1468 adjacency matrix representing connections between developers sharing at least one issue in common. To identify influential developers, we calculated four centrality measures: degree centrality (number of direct connections), betweenness centrality (frequency of appearing on shortest paths between nodes), closeness centrality (average shortest path length to all other nodes), and eigenvector centrality (influence based on connections to other influential nodes). Developers were ranked based on these measures to identify key contributors to sustainability discussions.

Next, we transformed the bi-adjacency matrix into a 2279 x 2279 adjacency matrix representing connections between issues. Each entry in this matrix indicated the number of developers who commented on both issues, resulting in a weighted and undirected network. This transformation was achieved using the dot product of the bi-adjacency matrix transpose.

For visualization, the network displayed issues as nodes and edges representing shared developer comments, with edge thickness proportional to the number of shared comments. We applied a color-coding scheme with 23 colors for specific topics (micro-labels) and 6 shapes for SusAF measures (macro-labels). Sixty isolated issues were removed to focus on the connected component.

Topic labels	Size	SusAF measures
Gas Price and Transaction Fees	355	Economic Environmental ●
Database State Management and Trie	321	Individual Technical ✚
P2P Network and Ethereum Build Process	319	Social Technical ✖
Account Security and Keystore Management	274	Individual Technical ✚
Swarm Network and Manifest Management	234	Social Technical ✖
Tracing and Debugging with Tracers	188	Technical ◆
Ethereum Source and Path Management	118	Technical ◆
Website Deployment and Management with Netlify	88	Technical ◆
Docker Image Management and Process Handling	84	Technical ◆
EVM (Ethereum Virtual Machine) and Stack Management	83	Technical ◆
Source and Path Management for Ethereum	80	Technical ◆
Light Ethereum Subprotocol (LES) Client Logic and Capacity	73	Technical ◆
Peer-to-Peer Protocol and Node Discovery	70	Technical ◆
Bootnode and Peer Management	69	Social Technical ✖
Peer Sync and Latency Management	66	Social Technical ✖
Repository Management and Issue Handling	52	Technical ◆
Whisper Protocol and Message Handling	48	Technical ◆
Framework and Library Management	46	Technical ◆
Benchmark Analysis and Optimization	46	Environmental Technical ▲
Blockchain Status and Transaction Logs	37	Technical ◆
Light Mode and Peer Management	36	Social Technical ✖
Crypto Benchmarking and Performance Analysis	33	Economic Technical ■
Community Support and Question Tracking	32	Social Technical ✖

TABLE III: Sustainability topics interpretation results. The symbols in the right column will be used later, in Sec. IV-C.

We applied the Leiden algorithm [43] to identify distinct communities within the network, revealing six distinct communities of issues. For each identified cluster, we calculated in-density (proportion of connections within a cluster) and out-density (proportion of connections between a cluster and the rest of the network). These densities were compared to those of randomly generated clusters by rewiring edges (hence producing 10^6 randomized versions of the real network) to assess the significance of the identified communities.

We then analyzed the distribution of SusAF measures (EC/EN, EC/T, EN/T, I/T, S/T, T) within each community to understand which sustainability topics were prevalent. Additionally, we examined the creation year of issues within each community to track the evolution of engagement with different sustainability topics over time.

IV. Results

In this section, we will discuss the findings derived from the topic extraction, the sustainability evaluation and the network analysis. These were conducted according to the guidelines detailed in Subsection III-C and within the framework established by Duboc et al. [44].

The Sustainability Awareness Framework (SusAF) delineates five dimensions of sustainability: economic, social, individual, environmental, and technical. The subsequent subsections will explore the topics identified within these five domains, detailing how the discussions among Go-Ethereum developers correspond to and support each sustainability dimension. In Tab. III, we outline the topics identified as relevant to sustainability by the model, with these findings later confirmed through manual verification. The column labelled "SusAF measure" indicates the specific dimensions of sustainability each topic pertains to, which we will explore in greater depth in the next section.

After setting aside issues classified as "Undefined," we found that 2830 out of the 8665 total issues are connected to sustainability, accounting for 32.66%.

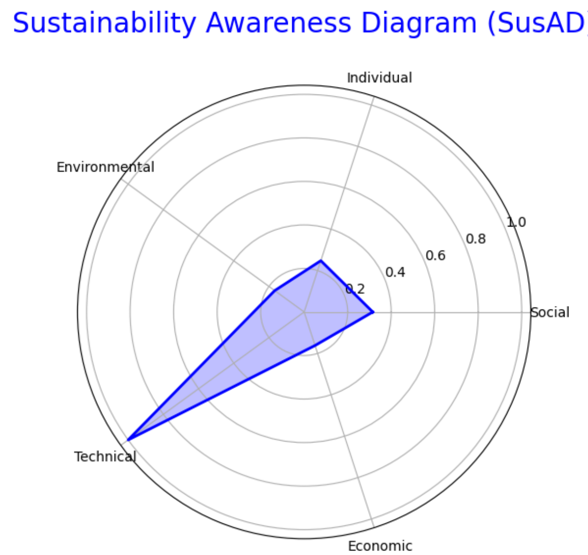


Fig. 5: Sustainability Awareness Diagram of the topic obtained

Figure 5 shows graphically how the discussions contribute to the sustainability of blockchain technology always considering the five measures defined above. The plot was obtained by considering the sum of topic sizes for each measure and then normalizing the data.

A. Sustainability evaluation

The dataset employed presents numerous topics of discussion among developers. This study aims to discern if any of these discussions address sustainability, gas and energy consumption, and the optimization of resource use. The concept of sustainability, as outlined by the United Nations Brundtland Commission in 1987, is to fulfill present needs without hindering future generations' capacity to meet theirs, a principle further emphasised in the 2030 Agenda for Sustainable Development[6].

This section begins with a broad analysis of the topics derived from our experiments. Subsequently, we examine the findings in light of the five sustainability dimensions defined by the Sustainability Awareness Framework (SusAF) [44], itself inspired by the Karlskrona Manifesto for Sustainability Design [45].

Sustainability is a key factor in blockchain design and development. The analysis of developer discussions, detailed in Table III, indicates that sustainability is either a primary or secondary focus in 22 out of 58 topics.

Figure 6 presents the top five topics as a percentage of all annual issues, highlighting the main subjects of conversation

[6]https://www.un.org/en/

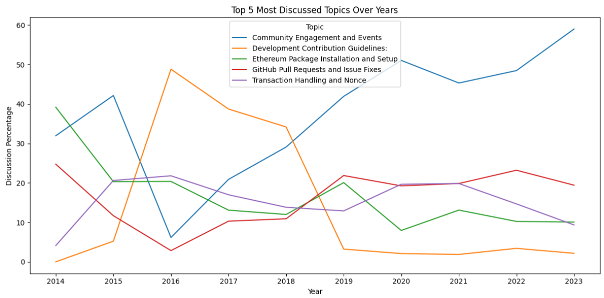


Fig. 6: Percentage of issues per topic over the annual amount

among developers, with a strong emphasis on the technical management of the platform.

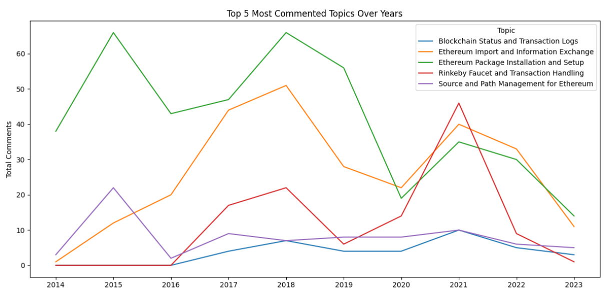


Fig. 7: Percentage of comments per topic over the annual amount

Figure 7 displays the top five topics based on their share of the total annual comments, reflecting the community's engagement with certain issues. Similar to the previous figure, the most commented-on topics by developers are technical in nature. For example, “Blockchain Status and Transaction Logs” has been considered as relevant to sustainability at the technical level by our framework. Conversations on these subjects contribute to overseeing blockchain functionality and the efficiency of transactions.

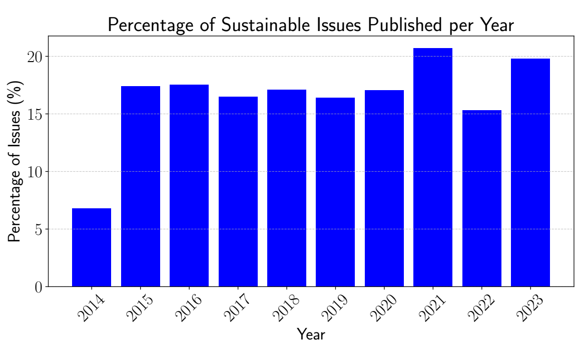


Fig. 8: Barplot of percentage of sustainable issues published over the total published each year

The graph in Figure 8 presents the ratio of discussions on sustainability-related topics to the overall number of topics tackled within the same year. The focus on sustainability among Ethereum developers reached its highest in 2021, with 2023 being a close second. This spike in discussion coincides with the Ethereum Merge on September 15, 2022, when Ethereum shifted from a proof-of-work to a proof-of-stake consensus mechanism, slashing its energy consumption by approximately 99.95%. This uptick in sustainability conversations likely occurred as the community prepared for the Merge by discussing its implementation and continued post-Merge to assess its impact on efficiency and ongoing operation. The dip in sustainability-related discussions in 2022, the year of the Merge, could be attributed to the predominance of technical issues related to the transition, which may have steered the community's focus away from broader sustainability topics.

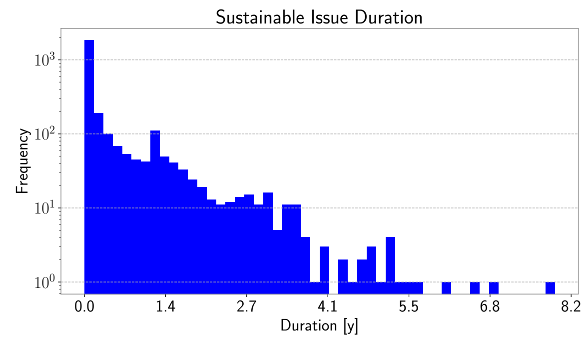


Fig. 9: Lifespan in years of sustainability related issues marked as in the dataset.

The mean lifespan of a sustainability issue, as we can see in the Figure 9 is 151 days instead of considering all the issues the mean lifespan is 128 days. So this highlights that sustainability issues in mean takes longer to resolve.

In our results we have 473 sustainable issues with 0 comments over an amount of 2752 sustainable issues. Over the sustainable issues discussion we have 1468 different users that comment an issue over 6694 users that comments all the topics issues that correspond a percentage of 21%. Of these 1468 commentators, very few are extremely active: only 8 participate in more than 100 issues, with an average of 3.8 issues they participate in per user.

Moreover, we analysed the number of comments for each sustainability issue, then we mapped each comment with the ID of the user that posted the comment for considering how many different users comment the issue over the total amount of comments. We find a correlation of 99.84% meaning that the users that comment are in general different. So the most commented issues coincide with the issues with the most commentators – meaning that the number of comments is a good proxy for issue popularity.

B. *SusAF measures discussion*

The Sustainability Awareness Framework (SusAF) outlines five dimensions of sustainability: economic, social, individual, environmental, and technical. The following subsections explore how Go-Ethereum developer discussions align with these dimensions.

1) Economic dimension: Economic sustainability focuses on value creation, customer relationships, supply chain efficiency, governance, and innovation. In the Go-Ethereum com-

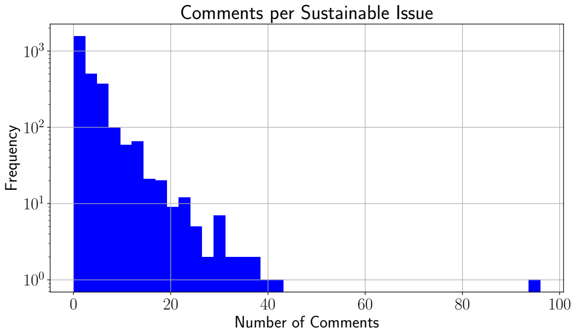


Fig. 10: Number of comments per sustainability related issues.

Fig. 11: Number of developers per issue.

munity, topics such as ”Gas Price and Transaction Fees” and ”Crypto Benchmarking and Performance Analysis” highlight economic sustainability [46]. Discussions emphasize reducing transaction fees and efficient network management to enhance adoption, user experience, and lower operational costs, thereby supporting blockchain’s economic sustainability [47].

2) Social dimension: Social sustainability emphasizes community cohesion, trust, inclusivity, equity, and active participation. Topics like ”P2P Network and Ethereum Build Process” and ”Swarm Network and Manifest Management” show the importance of community engagement and collaboration [48]. These discussions promote a robust, inclusive, and cooperative blockchain community, supporting fairness and diversity. Additionally, ”Community Support and Question Tracking” exemplifies social collaboration among developers to enhance blockchain usage [49].

3) Individual dimension: Individual sustainability includes health, lifelong learning, privacy, security, self-awareness, and free will. Topics like ”Database State Management and Trie” and ”Account Security and Keystore Management” highlight privacy, security, and empowerment [50]. These discussions promote innovative thinking, problem-solving, and data integrity, essential for individual sustainability.

4) Environmental dimension: Environmental sustainability focuses on resource management and reducing environmental impact. Discussions on ”Gas Price and Transaction Fees” and ”Benchmark Analysis and Optimization” address reducing energy demands and environmental footprint [51]. These topics emphasize waste minimization and efficient resource use,

User ID	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
129561	0.5062	0.3943	0.6430	0.4162
142290	0.3690	0.2318	0.5842	0.3270
6915	0.2564	0.1313	0.5441	0.2468
111600	0.1858	0.0853	0.5177	0.1631
10137	0.1809	0.0626	0.5168	0.1936

TABLE IV: Centrality measures of Go-Ethereum developers in all topics discussions.

User ID	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
129561	0.5488	0.4432	0.6838	0.4375
142290	0.4099	0.2758	0.6214	0.3383
6915	0.2801	0.1495	0.5745	0.2558
10137	0.2053	0.0850	0.5471	0.1866
111600	0.2029	0.1019	0.5460	0.1649

TABLE V: Centrality measures of Go-Ethereum developers in sustainability discussions.

supporting a sustainable relationship with natural resources.

5) Technical dimension: Technical sustainability involves maintaining, adjusting, securing, and scaling systems to adapt to changing environments. Issues like ”P2P Network and Ethereum Build Process,” ”Tracing and Debugging with Tracers,” and ”Repository Management and Issue Handling” demonstrate the importance of creating resilient and adaptable technological solutions [44]. These efforts ensure long-term sustainability and flexibility, fostering innovation and growth.

Finally, Table III shows how the extracted topics affect the five sustainability measures defined by Duboc et al. [44]. Developer discussions mainly contribute to the platform’s technical sustainability, with indirect benefits to social, individual, economic, and environmental aspects by reducing energy consumption and optimizing emissions.

To summarise our results, the first question, **RQ1**, asked: **Are sustainability issues a topic of discussion among Go-Ethereum developers?**

> **Answer to RQ1:** Our findings reveal that *out of the 58 topics identified, 22 were related to sustainability, indicating that sustainability is indeed a subject of discussion among Go-Ethereum developers*.

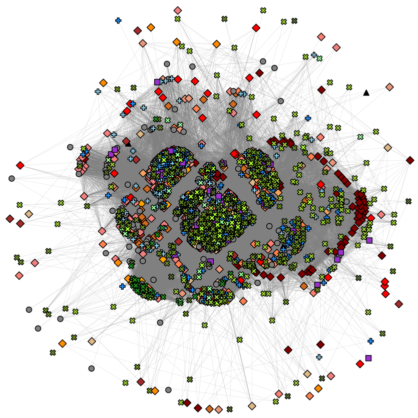
The second question, **RQ2**, asked: **How do sustainability-related discussions evolve over time?**

> **Answer to RQ2:** The analysis shows that *the proportion of sustainability-related issues peaked in 2021, closely followed by 2023, suggesting an increased focus on sustainability topics in recent years, particularly around the time of the Ethereum Merge event in September 2022.*

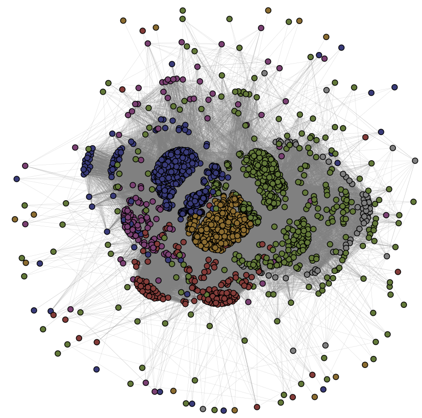
C. Sustainability Network Analysis

This section examines developer participation in sustainability discussions through a network analysis following the methodology explained in III-D. Nodes represent developers, and edges indicate shared comments on sustainability-related GitHub issues. The strength of connections increases with the number of shared comments, revealing collaboration and influence patterns.

We first consider the bipartite network developers-issues projection onto Layer 1 (developers only). To identify influ-



(a) Network of issues.



(b) Communities in network of issues.

Fig. 12: Left Panel: Network of commented sustainable issues. Links represent shared developer comments on issues, with thickness proportional to the number of shared comments. Nodes are colored by 23 sustainable topics and shaped by 6 SusAF measures (legend in Table III, right column).

Right Panel: Network colored by community detection using the Leiden algorithm, showing 6 communities: 649 (blue), 614 (green), 560 (yellow), 180 (red), 151 (purple), 65 (gray), totaling $2,219$ nodes.

ential developers, we used four centrality measures: **degree centrality**, which counts direct connections and shows how active a developer is; **betweenness centrality**, which identifies intermediaries who bridge groups; **closeness centrality**, which indicates how efficiently a developer can spread information; and **eigenvector centrality**, which highlights developers connected to other influential developers. Tables IV and V list the top five developers by user ID and centrality measure values. From Table V, we see that developer 129561 ranks highest across all centrality measures, highlighting their crucial role in sustainability discussions within the Go-Ethereum community. This developer is very active in discussions (**high degree centrality**), acts as a key information conduit (**high betweenness centrality**), is closely connected to others (**high closeness centrality**), and is influential among influential developers (**high eigenvector centrality**).
Developer 142290 also shows significant influence, particularly in connecting different groups and efficiently disseminating information. Developers 6915, 10137, and 111600 also rank highly across various metrics, underscoring their important roles in promoting sustainability discussions within the community. The third question, **RQ3**, asked: **Which sustainability-related developers are most influential? Are they also influential on the entire Go-Ethereum network?**

Answer to RQ3: Key developers facilitate collaboration and information exchange, potentially accelerating the adoption of sustainability practices within the Go-Ethereum project. The users crucial for sustainability are the same as those listed in Table IV, indicating their potential influence across the entire network.

We, then, consider the bipartite projection on Layer 2 (issues). Each node in the network represents an issue, and links represent shared comments by developers, with link thickness indicating the number of shared developers. We focused on the connected component of the network (i.e., the set of nodes connected by paths), excluding 60 isolated issues. Each issue is categorized under two labels: a *micro-label* for the specific topic (left column of Table III) and a *macro-label* for the related SusAF measure (right column of Table III). The dataset includes $2,279$ issues across 23 topics (micro-labels) classified into 6 SusAF measures (macro-labels): *EC/EN*, *EC/T*, *EN/T*, *I/T*, *S/T*, *T* (Economic & Environmental, Economic & Technical, Environmental & Technical, Individual & Technical, Social & Technical, Technical).

The distribution of macro-labels is: *EC/EN*, *EC/T*, and *EN/T* each with 1 topic; *I/T* with 2 topics; *S/T* with 6 topics; and *T* with 12 topics, totaling 23 topics.

We visualize the network with a color-coded scheme using 23 colors and 6 unique shapes for macro-labels, as detailed in Table III. The connected component of commented issues is shown in Fig. 12, Left Panel. The Leiden algorithm identifies six distinct communities, illustrated in different colors in Fig. 12, Right Panel [52].

Adopting the methodology described by Mungo et al. [19], we analyze a network of commented issues consisting of $N = 2219$ nodes and $k = 6$ clusters S, with $|S_i|$ representing the number of nodes in cluster S_i. We examine the in-density and out-density of links according to the partitioning established by the Leiden algorithm [43], [52]. Using the unweighted adjacency matrix A of the commented issues network and the clustering $S^* = \{S_1, \ldots, S_k\}$, the in-density for a cluster S_i is defined as:

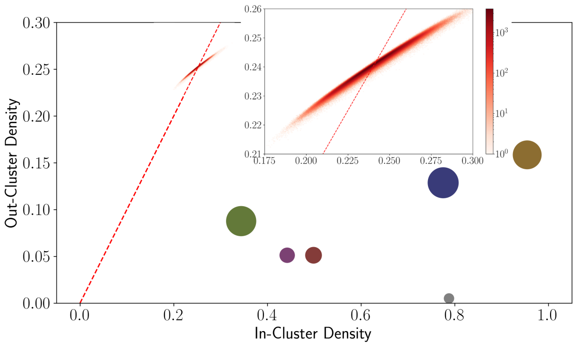


Fig. 13: Comparison of in- and out-cluster densities measured on the 5 clusters identified by the Leiden algorithm. Circles represent clusters from Fig. 12, Right Panel. Circle color corresponds to cluster, and size indicates cluster size. The dashed red line is the diagonal, with the red-shaded area showing in- and out-cluster density distribution for randomized networks. A zoom of the dashed area is shown at the top.

$$\rho_i^i = \frac{1}{|S_i|(|S_i|-1)} \sum_{j,k \in S_i, j \neq k} A_{jk},$$

and the out-density is given by:

$$\rho_i^o = \frac{1}{|S_i|(N-|S_i|)} \sum_{j \in S_i, k \notin S_i} A_{jk}.$$

Here, A_{jk} represents a binary value from the adjacency matrix indicating the presence of a link between nodes j and k, focusing solely on the existence of a connection between issues rather than its strength. We then compare the in-densities and out-densities of the clusters identified by the algorithm against those of randomly generated clusters. For the random clusters, each issue is assigned to one of the six possible clusters with equal probability; the simulation is obtained with $N = 10^6$ random networks.

We show the results in Fig. 13: the densities within clusters of similar issues (commented on by the same developers) are higher than those between different clusters. Comparing these results with random clusters (red shaded area) reveals significant differences, indicating that issues with similar features attract a common group of developers. This suggests developers focus on specific features rather than commenting across all issues.

Next, we categorize these features to identify common characteristics within clusters. Fig. 14 presents a bar plot showing the distribution of issues per SusAF dimensions (*EC/EN*, *EC/T*, *EN/T*, *I/T*, *S/T*, *T*) within each community. The x-axis ticks are color-coded as in Fig. 13 and Fig. 12, Right Panel. Each bar represents the number of issues of a specific topic within the clusters.

Community 0 (blue) and Community 2 (yellow) are mainly associated with *EC/EN* (Economic & Environmental) and *I/T* (Individual & Technical) topics. Community 1 (green) is distinct for *S/T* (Social & Technical). Technical (T) topics are dispersed throughout the network. The *EC/T* (Economic & Technical) and *EN/T* (Environmental & Technical) measures have only 33 and 46 issues, respectively, indicating limited relevance (Tab. III).

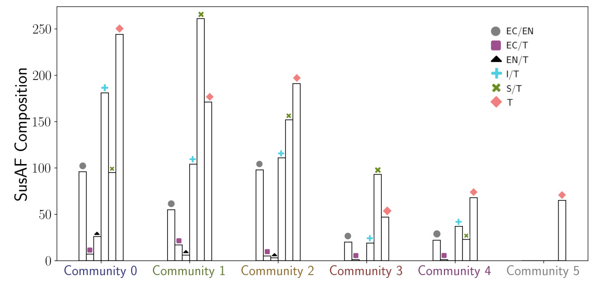


Fig. 14: Distribution of the SusAF measures across the six clusters. The x-axis shows the various communities, which are distinguished by the previously established colour coding. Within each community, the count of issues related to specific SusAF measures is shown, consistent with the legend provided in Tab. III.

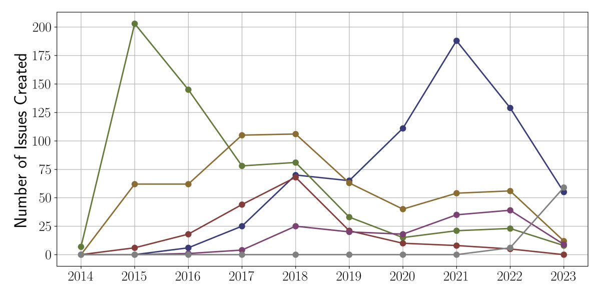


Fig. 15: Issues created over the years (2014-2023) per community. The communities are distinguished by the previously established colour coding.

We also analyze the creation year of the issues, presenting the temporal trends in Fig. 15.

We examine the temporal distribution of opened issues by community over the years, revealing distinct activity patterns. Community 1 (green, *S/T*) was most active from 2015 to 2017. Community 2 (yellow, *EC/EN* and *I/T*) saw increased activity during the middle years. Since 2020, Community 0 (blue, *EC/EN* and *I/T*) has shown a notable rise in engagement. Similar trends are observed when analyzing the last update dates, indicating consistent community engagement over time.

Therefore, we can conclude our network analysis by addressing **RQ4**: **Do developers engage in any topic discussed at the moment, or do they select topics based on their relative area (e.g., social rather than environmental)?**

Answer to RQ4: Our findings show that developers prefer engaging in discussions similar to their previous ones, rather than responding randomly. Community detection analysis reveals that developers consistently respond to Economic, Environmental, and Individual topics, with increased activity from 2019 onwards. Social issues were notably active from 2015 to 2017, indicating selective engagement. Technical issues are broadly discussed by all developers.

The communities identified by the Leiden algorithm, which include developers with shared interests, reflect issues that were initiated in the same year or in adjacent years. These findings are consistent with the SusAF categorization of topics we applied, effectively mirroring the thematic preferences and temporal engagement patterns within the developer community.

V. Threats to Validity

In conducting this study, we acknowledge several validity threats that could impact our findings and their interpretation:

External Validity: Our focus on GitHub discussions within the Go-Ethereum project may not fully represent the broader sustainability dialogue in the blockchain community, as discussions on other platforms or in private settings are not captured. This limits the generalizability of our findings. *Mitigation*: We chose one of the most prominent and active blockchain projects to provide valuable insights into industry trends.

Internal Validity: Our BERT-based method for topic extraction might miss subtle nuances, and the semi-supervised approach with manual validation introduces potential bias. Using the number of comments as a measure of issue discussion intensity might conflate prolonged issues with genuinely popular ones [53]. *Mitigation*: We used a multi-step validation process, including manual reviews and automated techniques. Future work should use additional metrics like the number of unique authors to better represent engagement.

Construct Validity: The SusAF framework might overlook emerging sustainability issues due to blockchain's rapid evolution. Limiting our analysis to discussions up until May 2023 risks relevance due to fast-paced developments. Interpretation bias from manual reviews and automated keyword extraction may affect replicability. *Mitigation*: We conducted regular literature reviews to stay updated on emerging issues. Our methodology and the code used is publicly available to address replicability concerns.

Our multi-method approach, combining topic modeling, network analysis, and manual validation, provides a robust foundation for our findings. Readers should consider these limitations when interpreting our results.

VI. Conclusion and future works

This study develops a method to analyze how Ethereum developers discuss sustainability on GitHub, enhancing our understanding of blockchain's environmental and efficiency concerns. Our findings show that while developers mainly focus on technical aspects, their discussions also cover economic, social, individual, environmental, and technical sustainability dimensions.

Our analysis reveals that sustainability is a significant topic among Ethereum developers, with 22 out of 58 topics related to it. Discussions have evolved, peaking around the Ethereum Merge in 2022, indicating increased awareness. Key developers driving these discussions are influential in the Go-Ethereum network, showing the integration of sustainability in the development process. Developers tend to engage with topics aligned with their interests, suggesting specialized knowledge clusters.

Our network analysis identifies distinct communities centered on specific sustainability dimensions, with gas prices and cost efficiency emerging as central themes. Economic considerations often serve as an entry point for broader sustainability discussions in blockchain development.

While insightful, our study is limited to GitHub discussions within a single project. Future research could expand to other blockchain projects and platforms for a broader view, incorporate quantitative sustainability measures like energy consumption data, and conduct longitudinal studies to track the evolution of sustainability considerations as blockchain technology matures.

References

[1] G. Tripathi, M. A. Ahad, and G. Casalino, "A comprehensive review of blockchain technology: Underlying principles and historical background with future challenges," *Decision Analytics Journal*, vol. 9, p. 100344, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772662223001844

[2] V. Kohli, S. Chakravarty, V. Chamola, K. S. Sangwan, and S. Zeadally, "An analysis of energy consumption and carbon footprints of cryptocurrencies and possible solutions," *Digital Communications and Networks*, vol. 9, no. 1, pp. 79–89, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352864822001390

[3] A. Deshpande, K. Stewart, L. Lepetit, and S. Gunashekar, "Distributed ledger technologies/blockchain: Challenges, opportunities and the prospects for standards," *Overview report The British Standards Institution (BSI)*, vol. 40, p. 40, 2017.

[4] E. Ghosh and B. Das, "A study on the issue of blockchain's energy consumption," in *Proceedings of International Ethical Hacking Conference 2019*, M. Chakraborty, S. Chakrabarti, and V. E. Balas, Eds. Singapore: Springer Singapore, 2020, pp. 63–75.

[5] M. R. Islam, M. M. Rashid, M. A. Rahman, M. H. S. B. Mohamad *et al.*, "A comprehensive analysis of blockchain-based cryptocurrency mining impact on energy consumption," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[6] M. Vaccargiu, S. Aufiero, S. Bartolucci, R. Neykova, R. Tonelli, and G. Destefanis, "Sustainability in blockchain development: A bert-based analysis of ethereum developer discussions," in *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*, 2024, pp. 381–386.

[7] K. Qin and A. Gervais, "An overview of blockchain scalability, interoperability and sustainability," *Hochschule Luzern Imperial College London Liquidity Network*, pp. 1–15, 2018.

[8] A. Arshad, F. Shahzad, I. Ur Rehman, and B. S. Sergi, "A systematic literature review of blockchain technology and environmental sustainability: Status quo and future research," *International Review of Economics Finance*, vol. 88, pp. 1602–1622, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105905602300254X

[9] R. Asif and S. R. Hassan, "Shaping the future of ethereum: Exploring energy consumption in proof-of-work and proof-of-stake consensus," *Frontiers in Blockchain*, vol. 6, p. 1151724, 2023.

[10] A. Pinna, G. Baralla, M. Marchesi, and R. Tonelli, "Raising sustainability awareness in agile blockchain-oriented software engineering," in *2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2021, pp. 696–700.

[11] N. Eligüzel, "An analysis of the integration of sustainability concepts into blockchain technology," *International Journal of Applied Methods in Electronics and Computers*, vol. 11, no. 3, p. 158–164, Sep. 2023. [Online]. Available: https://www.ijamec.org/index.php/ijamec/article/view/382

[12] Y. Liu, S. Zhang, M. Chen, Y. Wu, and Z. Chen, "The sustainable development of financial topic detection and trend prediction by data mining," *Sustainability*, vol. 13, no. 14, 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/14/7585

[13] E. Stamoulis, "Comparative study on the environmental, political, social effects and long-term sustainability of bitcoin, ethereum, tether and cardano cryptocurrencies," August 2021. [Online]. Available: http://essay.utwente.nl/88266/

[14] A. Ayman, S. Roy, A. Alipour, and A. Laszka, "Smart contract development from the perspective of developers: Topics and issues discussed on social media," in *Financial Cryptography and Data Security: FC 2020 International Workshops, AsiaUSEC, CoDeFi, VOTING, and WTSC, Kota Kinabalu, Malaysia, February 14, 2020, Revised Selected Papers 24*. Springer, 2020, pp. 405–422.

[15] G. Ibba and M. Vaccargiu, "Analysis of users' most discussed topics and trends on blockchain technologies and smart contracts," in *2023 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2023, pp. 865–873.

[16] B. T.-S. Kim and E.-J. Hyun, "Mapping the landscape of blockchain technology knowledge: A patent co-citation and semantic similarity approach," *Systems*, vol. 11, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2079-8954/11/3/111

[17] A. Briola and T. Aste, "Dependency structures in cryptocurrency market from high to low frequency," *Entropy*, vol. 24, no. 11, p. 1548, 2022.

[18] D. Vidal-Tomás and S. Bartolucci, "Artificial intelligence and digital economy: divergent realities," *Available at SSRN 4589333*, 2023.

[19] L. Mungo, S. Bartolucci, and L. Alessandretti, "Cryptocurrency co-investment network: token returns reflect investment patterns," *EPJ Data Science*, vol. 13, no. 1, p. 11, 2024.

[20] D. Lin, J. Wu, Q. Yuan, and Z. Zheng, "Modeling and understanding ethereum transaction records via a complex network approach," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2737–2741, 2020.

[21] S. Aufiero, G. Ibba, S. Bartolucci, G. Destefanis, R. Neykova, and M. Ortu, "Dapps ecosystems: mapping the network structure of smart contract interactions," *EPJ Data Science*, vol. 13, no. 1, 2024, all Open Access, Gold Open Access. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85205696190&doi=10.1140%2fepjds%2fs13688-024-00497-8&partnerID=40&md5=115ded03111e24ef22b53f12a0650e59

[22] G. Destefanis, "Complex systems oriented approach for dapps analysis," in *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2024, pp. 757–762.

[23] G. Ibba, S. Aufiero, S. Bartolucci, R. Neykova, M. Ortu, R. Tonelli, and G. Destefanis, "Mindthedapp: a toolchain for complex network-driven structural analysis of ethereum-based decentralised applications," *IEEE Access*, 2024.

[24] G. Ibba, S. Aufiero, R. Neykova, S. Bartolucci, M. Ortu, R. Tonelli, and G. Destefanis, "A curated solidity smart contracts repository of metrics and vulnerability," in *Proceedings of the 20th International Conference on Predictive Models and Data Analytics in Software Engineering*, 2024, pp. 32–41.

[25] G. Ibba, S. Khullar, E. Tesfai, R. Neykova, S. Aufiero, M. Ortu, S. Bartolucci, and G. Destefanis, "A preliminary analysis of software metrics in decentralised applications," in *Proceedings of the Fifth ACM International Workshop on Blockchain-enabled Networked Sensor Systems*, 2023, pp. 27–33.

[26] G. Ibba, G. Destefanis, R. Neykova, M. Ortu, S. Aufiero, and S. Bartolucci, "Dai: A dependencies analyzer and installer for solidity smart contracts," in *2024 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE*, 2024.

[27] S. Bartolucci, G. Destefanis, M. Ortu, N. Uras, M. Marchesi, and R. Tonelli, "The butterfly "affect": Impact of development practices on cryptocurrency prices," *EPJ Data Science*, vol. 9, no. 1, p. 21, 2020.

[28] L. Lucchini, L. Alessandretti, B. Lepri, A. Gallo, and A. Baronchelli, "From code to market: Network of developers and correlated returns of cryptocurrencies," *Science advances*, vol. 6, no. 51, p. eabd2204, 2020.

[29] M. Ortu, N. Uras, C. Conversano, S. Bartolucci, and G. Destefanis, "On technical trading and social media indicators for cryptocurrency price classification through deep learning," *Expert Systems with Applications*, vol. 198, p. 116804, 2022.

[30] H. Mumtaz, C. Paradis, F. Palomba, D. A. Tamburri, R. Kazman, and K. Blincoe, "A preliminary study on the assignment of github issues to issue commenters and the relationship with social smells," *Proceedings - 15th International Conference on Cooperative and Human Aspects of Software Engineering, CHASE 2022*, p. 61 – 65, 2022. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85133979998&doi=10.1145%2f3528579.3529181&partnerID=40&md5=469de3e90f0e156527320fb5018b7aec

[31] J. Jamieson, N. Yamashita, and E. Foong, "Predicting open source contributor turnover from value-related discussions: An analysis of github issues," in *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE)*, 2024, pp. 678–690.

[32] M. Ortu, T. Hall, M. Marchesi, R. Tonelli, D. Bowes, and G. Destefanis, "Mining communication patterns in software development: A github analysis," in *Proceedings of the 14th international conference on predictive models and data analytics in software engineering*, 2018, pp. 70–79.

[33] M. Ortu, G. Destefanis, T. Hall, and D. Bowes, "Fault-insertion and fault-fixing behavioural patterns in apache software foundation projects," *Information and Software Technology*, vol. 158, p. 107187, 2023.

[34] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[35] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: Tf-idf approach," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 61–66.

[36] U. M. R. M. A. A. A. W. A. J. A. S. Khan MQ, Shahid A, "Impact analysis of keyword extraction using contextual word embedding," *PeerJ Computer Science*, 2022. [Online]. Available: https://peerj.com/articles/cs-967/

[37] T. Ma, J.-G. Yao, C.-Y. Lin, and T. Zhao, "Issues with entailment-based zero-shot text classification," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 786–796. [Online]. Available: https://aclanthology.org/2021.acl-short.99

[38] P. R. Nair and D. R. Dorai, "Evaluation of performance and security of proof of work and proof of stake using blockchain," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, pp. 279–283.

[39] S. Blanco-Cuaresma, I. Ciucă, A. Accomazzi, M. J. Kurtz, E. A. Henneken, K. E. Lockhart, F. Grezes, T. Allen, G. Shapurian, C. S. Grant, D. M. Thompson, T. W. Hostetler, M. R. Templeton, S. Chen, J. Koch, T. Jacovich, D. Chivvis, F. de Macedo Alves, J.-C. Paquin, J. Bartlett, M. Polimera, and S. Jarmak, "Experimenting with large language models and vector embeddings in nasa scix," 2023.

[40] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the eighth ACM international conference on Web search and data mining*, 2015, pp. 399–408.

[41] E. Rijcken, F. Scheepers, K. Zervanou, M. Spruit, P. Mosteiro, and U. Kaymak, "Towards interpreting topic models with chatgpt," in *The 20th World Congress of the International Fuzzy Systems Association*, 2023.

[42] G. Colavito, F. Lanubile, N. Novielli, and L. Quaranta, "Leveraging gpt-like llms to automate issue labeling," in *2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)*. IEEE, 2024, pp. 469–480.

[43] V. A. Traag, L. Waltman, and N. J. Van Eck, "From louvain to leiden: guaranteeing well-connected communities," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[44] L. Duboc, B. Penzenstadler, J. Porras, S. Akinli Kocak, S. Betz, R. Chitchyan, O. Leifler, N. Seyff, and C. C. Venters, "Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow," *Requirements Engineering*, vol. 25, pp. 469–492, 2020.

[45] C. Becker, R. Chitchyan, L. Duboc, S. Easterbrook, B. Penzenstadler, N. Seyff, and C. C. Venters, "Sustainability design and software: The karlskrona manifesto," in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 2. IEEE, 2015, pp. 467–476.

[46] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Khan, "A review of blockchain technology applications for financial services," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 3, p. 100073, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772485922000606

[47] "Hybridizing cost saving with trust for blockchain technology adoption by financial institutions," *Telematics and Informatics Reports*, vol. 6,

p. 100008, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772503022000068

[48] "Conceptualising a social business blockchain: The coexistence of social and economic logics," *Technological Forecasting and Social Change*, vol. 172, p. 120997, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162521004297

[49] M. Kassen, "Understanding decentralized civic engagement: Focus on peer-to-peer and blockchain-driven perspectives on e-participation," *Technology in Society*, vol. 66, p. 101650, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X21001251

[50] J. Cha, S. K. Singh, Y. Pan, and J. H. Park, "Blockchain-based cyber threat intelligence system architecture for sustainable computing," *Sustainability*, vol. 12, no. 16, p. 6401, 2020.

[51] A. Parmentola, A. Petrillo, I. Tutore, and F. De Felice, "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*, vol. 31, no. 1, pp. 194–217, 2022.

[52] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[53] S. K. Basak, L. Neil, B. Reaves, and L. Williams, "What challenges do developers face about checked-in secrets in software artifacts?" in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2023, pp. 1635–1647.