Substitute and Complementary Open Source Software in Blockchain

S. Iwami[1], Y. Takefuji[2]
[1]OSS Promotion Center, NEC Corporation, Tokyo, Japan
[2]Faculty of Data Science, Musashino University, Tokyo, Japan
(iwami.research@outlook.com)

2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) | 979-8-3503-2315-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/IEEM58616.2023.10406790

Abstract - **Recently, blockchain technologies seem to have emerged from a period of disillusionment named in the hype cycle, and development has become active again. In this research, substitute and complementary repositories were identified from GitHub records in order to grasp the state of representative blockchain platforms: Bitcoin, Ethereum, Hyperledger, Ripple, and Corda. Within many blockchain platforms, it is common to have complementary relationships. Ethereum and Hyperledger also have a complementary relationship across platforms. The results showed that Ethereum is reactivating Hyperledger, whose development is stabilizing. This research proposes a methodology to find next-step software development from the network based on developers' skills via their movements between repositories.**

Keywords - **OSPO, open source program office, development, blockchain, promising technology**

I. INTRODUCTION

Watching topic shifts of co-occurring keyword networks, the cluster rank about blockchain recovered in 2023 after it downgraded in 2020-2022, as shown in Table I [1] analyzed as of 2023-04-27. Around 2017, the system concepts with blockchain became popular among information technology (IT) engineers. One reason for the fall might be that traditional databases had enough ability to meet our demands when we built a system.

TABLE I
Ranks and Top 5 Keywords of Clusters about Blockchain

Period	Rank by Clustering Method	
	Louvain	Infomap
2014-2016	12 (video; blockchain; p2p; streaming; webrtc)	25 (privacy; encryption; cryptography; p2p; blockchain)
2017-2019	13 (blockchain; ethereum; cryptography; p2p; bitcoin)	18 (blockchain; ethereum; cryptography; bitcoin; cryptocurrency)
2020-2022	14 (blockchain; ethereum; cryptography; cryptocurrency; encryption)	41 (blockchain; ethereum; solidity; smart-contracts; evm)
2023	11 (video; blockchain; ethereum; audio; cryptography)	22 (blockchain; ethereum; cryptocurrency; bitcoin; crypto)

From the transition of keywords In Table I, the period 2020-2022 seems to have been thinking about blockchain applications such as smart-contracts, but the year 2023 seems to return to the same keyword trends as before 2019. COVID-19 (coronavirus disease 2019) might impact these trends. It should be noted that 2023 reflects only four months of data, so trends may change in the future. If such a change is detected, it is necessary for the Open Source Program Office (OSPO) staff who oversee the organization's open source software (OSS) utilization to conduct a specific survey of OSS in the relevant field and grasp the current situation.

A. Role of OSPO and Necessity of this Analysis

The utilization of open source has become a matter of course, and open source has become indispensable in business that uses IT. Even if it is easy to use open source for personal use, it is necessary to overcome issues such as licensing and security measures to use it safely in business. Therefore, the OSPO is the department that oversees OSS within the organization.

Some OSPOs work in collaboration with the chief technology officer (CTO), who is deeply involved in technical strategy planning. Regardless of open source, technology strategy, in general, is also leveraging data to make better decisions. Then, a similar trend started to come to OSPO in some organizations, for example, Google, Red Hat [2], Sony, and NEC Corporation [1], according to the contributions to the events such as Mining Software Repositories and Open Source Summit. OSPO welcomes the automatic extraction of OSS trends and signs of changes through data science.

B. Research Purpose

The purpose of this research is to identify the promising OSS deeply related to the ongoing developments of OSS. Identifying the next OSS to be developed through the movement of contributors has two benefits: leveraging expert knowledge and reusing skills. The results will be beneficial for decision-makers in OSPOs to build their OSS strategy based on evidence.

C. Literature Review

The substitute goods, complementary goods, and independent goods are from a theory taught in microeconomics [3], and it is available to build a strategy [4]. In the engineering section, complementary technologies have revived declining technologies [5], and it has been considered to sell own products as complementary goods of OSS [6]. One previous study identified substitute and complementary products for assortment optimization. It applied machine learning algorithms for the data of transactions [7]. Another previous study simulated substitute and complementary

robots [8]. Substitution and complementary OSS of this methodology were detected in another dataset about AI framework [9].

The rest of this article is organized as follows: Section 2 describes the methodology of this research. Section 3 provides results, and section 4 discusses the results. Section 5 summarizes this research.

II. METHODOLOGY

In brief, the analysis is performed by the following steps:

(1) GitHub repositories about blockchain are selected as target repositories.
(2) Contributor accounts are listed from the target repositories.
(3) Repositories contributed by the listed contributor accounts are extracted. These repositories are named related repositories in this research.
(4) Shared rate of contributions and correlations coefficient of yearly numbers of contributors between a target repository and the extracted repositories are computed.

Related repositories are limited by the conditions: the number of stars is 1,000 or more, and the number of shared contributors is ten or more.

A. Target Repositories

The following repositories, whose source code is managed on GitHub, were selected as targets.

- https://github.com/bitcoin/bitcoin
- https://github.com/ethereum/go-ethereum
- https://github.com/hyperledger/fabric
- https://github.com/XRPLF/rippled
- https://github.com/corda/corda

Blockchain was proposed in 2008 [9], and the originator developed Bitcoin in 2009 [10]. Ripple was a transaction protocol and started to work as a cryptocurrency in 2013 [11]. Ethereum was published in 2014 [12]. Hyperledger was started in 2015 by the Linux Foundation [13]. Corda was launched in 2016 [14]. These blockchain platforms are competing in the industry, especially in the financial sector.

B. Indicators

Several indicators are computed as follows:

- Stars - The number of stars is extracted from the stars via GitHub API. This indicator is used to decide the priority of established OSS repositories.
- Slope - The value of the slope is calculated from the slope of a linear function. In other words, it is the value a of the equation "y = ax + b" by the least-square method for linear functions. When it is calculated, the data of the last year is removed because it is growing in the year. This indicator is used to decide the priority of emerging OSS repositories.
- Share - The share of contributors between a target repository and its related repository is defined as the rate of the number of contributors who attend the target repository and the related repository. In other words, when drawing a Venn diagram with the target repository and the related repository, the share is that the number of AND area is divided by the number of OR area as shown in Fig. 1.
- Correlation Coefficient (CC) - The correlation coefficient is computed with the function "corrcoef" of the numpy Python module [15]. The function output a slope and an intercept of the vertical axis in the equation of the liner function. The input values are the two timelines of contributors of the target and related repositories. The timelines are prepared from the number of contributors committing to the repository each year.

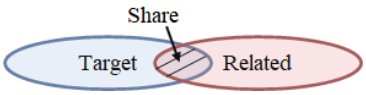

Fig. 1. Position of Share.

III. RESULTS

This section presents OSS repositories that may be the substitute and complementary as a result of the analysis. However, regarding Corda, its related repositories via contributors are not extracted. Thus, the results of the rest four blockchain platforms are shown.

The scores of the target repositories are tabulated in Table II. Ripple started earlier than Ethereum and Hyperledger, but Ripple gained fewer stars than Ethereum and Hyperledger. The Hyperledger Fabric is slightly declining based on the negative slope, but a lot of Hyperledger derivatives have been developed.

TABLE II
Stars and Slope of Target OSS Repositories

URL of Targets	Stars	Slope
https://github.com/bitcoin/bitcoin	69295	13.45274725
https://github.com/ethereum/go-ethereum	42042	18.25454545
https://github.com/hyperledger/fabric	14531	-0.892857143
https://github.com/XRPLF/rippled	4302	1.821678322
https://github.com/corda/corda	3926	2.94047619

To judge if the correlation coefficient is causal, the rate of contributors shared between the target repositories and related repositories is shown, but the threshold or how to set the threshold is not yet revealed. Fig. 2 shows the distribution and percentile of the share of contributors. Although it is not possible to determine the threshold of

share yet, based on the data distribution, this manuscript defines a high share as 0.1 or more.

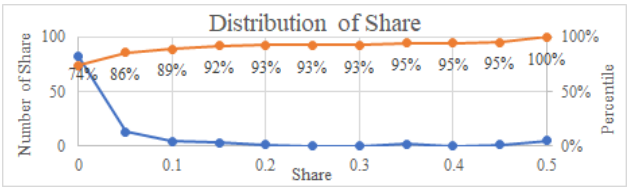

Fig. 2. Distribution of Share of Contributors between Repositories Related to Blockchains.

A. Bitcoin

The higher the share, the more valuable the CC is to see. Among 47 repositories, 38 data with positive slopes are plotted in Fig. 3. In Table III, the repositories in the high share group and the top repositories about stars and slope are shown. Among the three repositories with a large share, two with a very high CC (**Red Bold**) are related to Bitcoin, as judged from the path, and one with a high CC (*Blue Italic*), zcash which was forked from Bitcoin, shares contributors with Bitcoin. However, the slope has not gone beyond Bitcoin itself.

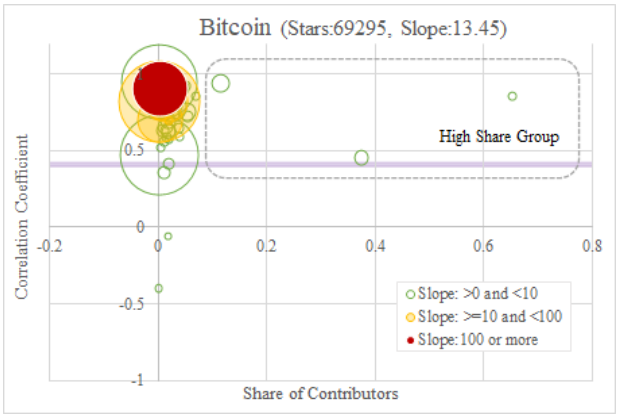

Fig. 3. Share, Correlation Coefficients, Stars (size), and Slope (color) of Bitcoin-related Repositories.

TABLE III
Correlation Coefficient with Bitcoin.

Related OSS	Stars	Slope	Share	CC
/Bitcoin-ABC/bitcoin-abc	1120	6.5407	0.6517	**0.8570**
/zcash/zcash	4730	0.7121	0.3747	*0.4539*
/bitcoin/bips	7958	5.0839	0.1150	**0.9418**
Reference: The Largest Number of Stars				
/tensorflow/tensorflow	173836	60.4762	0.0028	**0.8188**
Reference: The Largest Slope				
/DefinitelyTyped/DefinitelyTyped	43799	319.8	0.0016	**0.8752**

Searching other blockchain platforms for Bitcoin results, none of the other repositories involved Ethereum, Hyperledger, Ripple, and Corda. Thus, Bitcoin is independent of the other platforms, and Bitcoin has three complementary OSS inside the Bitcoin platforms and a Bitcoin-oriented repository.

B. Ethereum

Regarding Ethereum of 41 repositories, 36 data with positive slopes are plotted in Fig. 4. Table IV lists the repositories in the high share group, and the top repositories about stars and slope are shown. Among the two repositories with a large share, one with a very high CC (**Red Bold**), Erigon, is an Ethereum client, and another with a high CC (*Blue Italic*), Quorum, an Ethereum-based distributed ledger protocol.

Searching other blockchain platforms for Ethereum results, repositories related to Hyperledger are found: Hyperledger Fabric (share: 0.0120, CC: 0.8253) and Hyperledger Besu (share: 0.0180, CC: 0.6949). Shares between 0.01 and 0.02 are in the threshold band where we cannot tell whether it is causation or pseudo-correlation. Thus, it cannot be asserted that there is a relationship between substitute goods or complementary goods with other blockchain platforms, but since there are efforts for blockchain interoperability [16], there may be demands for participation in both platforms.

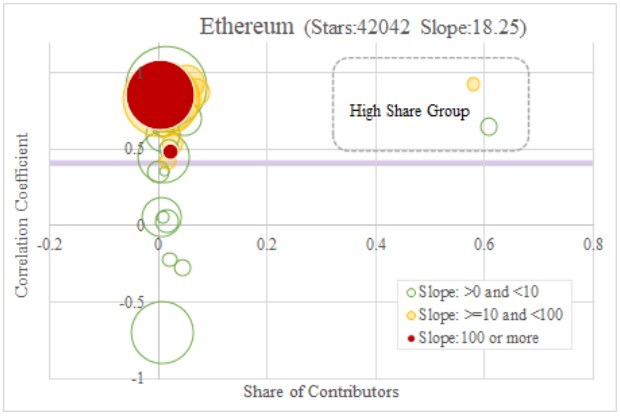

Fig. 4. Share, Correlation Coefficients, Stars (size), and Slope (color) of Ethereum-related Repositories.

TABLE IV
Correlation Coefficient with Ethereum.

Related OSS	Stars	Slope	Share	CC
/ConsenSys/quorum	4324	5.0788	0.6073	*0.6481*
/ledgerwatch/erigon	2336	12.5818	0.5794	**0.9288**
Reference: The Largest Number of Stars				
/golang/go	110804	7.0885	0.0152	**0.9055**
Reference: The Largest Slope				
/ethereum/ethereum-org-website	4020	142.9000	0.0221	*0.4846*

C. Hyperledger Fabric

The Hyperledger (/hyperledger/hyperledger) itself was already archived, so the data on GitHub were not gained. Instead, the data of Hyperledger Fabric, the most used platform in the Hyperledger-named platforms, is analyzed as a representative of Hyperledger group. Regarding Hyperledger Fabric of six repositories, three data with positive slopes are plotted in Fig. 5. The collection of samples of Hyperledger (/hyperledger/fabric-samples) has a high share with Hyperledger Fabric, and it

rises and falls together. As seen in the previous section, there are contributors participating in both Hyperledger Fabric and Ethereum.

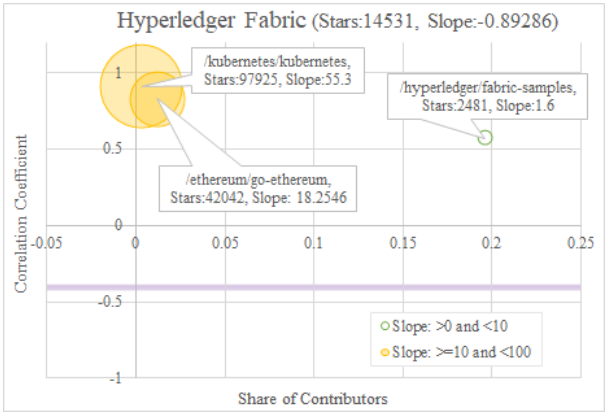

Fig. 5. Share, Correlation Coefficients, Stars (size), and Slope (color) of Hyperledger Fabric-related Repositories.

Hyperledger Fabric is popular in the industry, as shown by the number of stars, but in this analysis, there are few repositories extracted through shared contributors, and the slope is also a negative value. Therefore, among the frameworks named Hyperledger, repositories that can output results with 200 or more stars are additionally analyzed in Fig. 6. Data is as of 2023-05-02. These are shown in Table V.

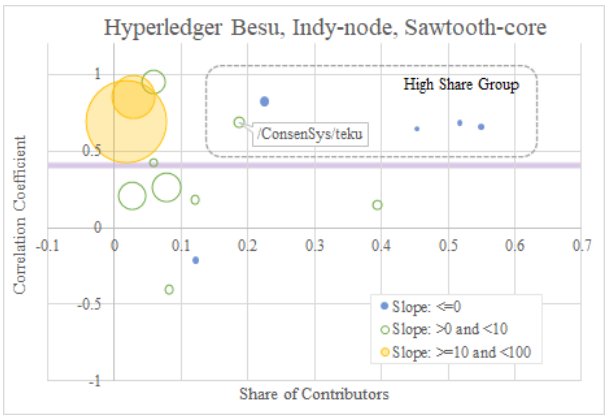

Fig. 6. Share, Correlation Coefficients, Stars (size), and Slope (color) of Repositories of Hyperledger Group.

TABLE V
Stars and Slope of Hyperledger Group.

URL of Targets	Stars	Slope
https://github.com/hyperledger/besu	1142	4.5
https://github.com/hyperledger/indy-node	642	-1.071428571
https://github.com/hyperledger/sawtooth-core	1416	-4.285714286

Hyperledger Indy and Sawtooth have a negative slope, like Hyperledger Fabric. A negative slope was also plotted in Fig. 6 in hopes of finding the relationship of substitute OSS, but no substitute OSS was found.

The only finding is Teku (*/ConsenSys/teku*; Share: 0.1862, CC: 0.6896), which is complementary OSS to Hyperledger Besu. ConsenSys, the owner of Teku, also appeared as the owner of Quorum in the Ethereum section. Therefore, it is considered that Hyperledger Besu and Teku are complementary OSS.

Overall, the development of Hyperledger tends to be inactive, but Hyperledger-related repositories related to Ethereum are active.

D. Ripple

Regarding Ripple of three repositories, two data with positive slopes are plotted in Fig. 7. Although the number of results is small, Ripple is independent of the other platforms, and Ripple has two complementary OSS inside the Ripple platform.

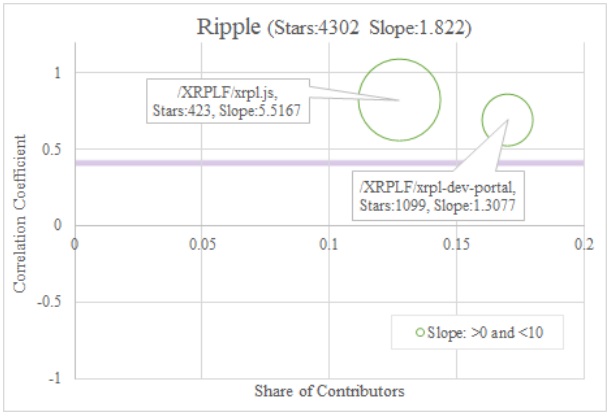

Fig. 7. Share, Correlation Coefficients, Stars (size), and Slope (color) of Ripple-related Repositories.

IV. DISCUSSION

A. Contribution

If someone is an expert in that technology field, she/he may be able to stay up to date with industry news and industry events and finds technology replacements and partnerships. However, such information targets major technologies, and there are many hidden technologies behind them.

This analysis helps prevent oversights, gives evidence to experts, and supports non-experts in catching the signs.

B. Limitation

Due to the intentional time-consuming data collection for preventing traffic rash, this analysis cannot provide the results by daily granularity. In addition, it will take more than one month to go through all the data. The system built in this research has made it possible to set the preferred repository in the source code, but some delays are inevitable.

In order to prevent an explosive increase of data, the setting of the system can limit data based on the number of stars and the number of shared contributors. Thus, very

small repositories were excluded from the analysis.

In this research, CC was computed on an annual timeline, but if there is a change during the year, it is possible that the signs will not be captured at the end of the year when the data increase is in progress. It is hoped that future updates will use monthly data.

V. CONCLUSION

This research searched the next-step OSS in the blockchain by means of identifying substitute goods and complementary goods. Consequently, some pieces of complementary OSS were found about Bitcoin, Ethereum, Hyperledger, and Ripple. In this analysis, trends in each blockchain are considered on the basis of the top repositories by share. However, for practical users such as enterprises, longer and wider lists will be required to increase options matching the user's talents and strategies.

Bitcoin and Ripple are closed inside each repository, and they have complementary OSS. In other words, they tend to be independent of other blockchain projects. Ethereum and Hyperledger Fabric have shared contributors, and it can be said that they are collaborating a bit. The growing trends of Ethereum-related repositories are stronger than the declining trends of Hyperledger-related repositories. This will be a matter of interoperability. One possibility is that interoperability development may be in high demand for other blockchain platforms as well.

The analysis of Hyperledger showed cases in which neither the owner's name (the project on GitHub) nor the contributors were shared, though it was under the umbrella of the Hyperledger Foundation. This should be kept in mind when the methodology of this analysis is automated and widely applied. However, it does not preclude the usefulness of this analysis.

In future works, it will be pursued to identify the threshold of share for causality. Potentially, it would be considered strong causality between repositories with the same owner, which is [owner] of */[owner]/[repository]* in a uniform resource locator (URL).

REFERENCES

[1] S. Iwami, "A Methodology to Categorize Rapidly Changing Projects of Open Source Software," in *2022 Portland International Conference on Management of Engineering and Technology (PICMET)*, 2022, pp. 1–8.

[2] D. Mueller and D. Izquierdo-Cortazar, "From Art to Science: The Evolution of Community Development," *IEEE Softw.*, vol. 36, no. 6, pp. 23–28, Nov. 2019.

[3] P. Newman, "Substitutes and Complements," in *The New Palgrave Dictionary of Economics*, London: Palgrave Macmillan UK, 1987, pp. 1–7.

[4] E. Han, "COMPLEMENTS VS. SUBSTITUTES: WHAT'S THE DIFFERENCE?," 2022. [Online]. Available: https://online.hbs.edu/blog/post/complements-vs-substitutes. [Accessed: 04-May-2023].

[5] S. Iwami, A. Ojala, C. Watanabe, and P. Neittaanmäki, "A bibliometric approach to finding fields that co-evolved with information technology," *Scientometrics*, vol. 122, no. 1, pp. 3–21, Jan. 2020.

[6] E. Haruvy, S. P. Sethi, and J. Zhou, "Open Source Development with a Commercial Complementary Product or Service," *Prod. Oper. Manag.*, vol. 17, no. 1, pp. 29–43, Jan. 2008.

[7] S. Tkachuk, A. Wroblewska, J. Dabrowski, and S. Lukasik, "Identifying Substitute and Complementary Products for Assortment Optimization with Cleora Embeddings," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–7.

[8] F. V. Abramov, "Implementation of an Adapted Ant Algorithm in the Presence of Substitute and Complementary Resources. Modeling the Behavior of the Manufacturer," in *2021 IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek)*, 2021, pp. 341–345.

[9] S. Iwami and Y. Takefuji, "Substitute and Complementary Open Source Software in Artificial Intelligence Frameworks," 17th European Conference on Software Architecture (ECSA), 2023. [Online]. Available: https://conf.researchr.org/details/ecsa-2023/ecsa-2023-industry-program/8/Substitute-and-Complementary-Open-Source-Software-in-Artificial-Intelligence-Framewor. [Accessed: 30-Sep-2023].

[10] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf. [Accessed: 02-May-2023].

[11] S. Nakamoto, "Bitcoin v0.1 released," 2009. [Online]. Available: https://www.mail-archive.com/cryptography@metzdowd.com/msg10142.html. [Accessed: 10-Jan-2021].

[12] M. E. Peck, "Ripple Credit System Could Help or Harm Bitcoin," *IEEE Spectrum*, 2013. [Online]. Available: https://spectrum.ieee.org/ripple-credit-system-could-help-or-harm-bitcoin. [Accessed: 02-May-2023].

[13] V. Buterin, "ETHEREUM: A NEXT-GENERATION CRYPTOCURRENCY AND DECENTRALIZED APPLICATION PLATFORM," *Bitcoin Magazine*, 2014. [Online]. Available: https://bitcoinmagazine.com/business/ethereum-next-generation-cryptocurrency-decentralized-application-platform-1390528211. [Accessed: 02-May-2023].

[14] The Linux Foundation and The Hyperledger Foundation, "Linux Foundation Unites Industry Leaders to Advance Blockchain Technology," 2015. [Online]. Available: https://www.hyperledger.org/announcements/2015/12/17/linux-foundation-unites-industry-leaders-to-advance-blockchain-technology. [Accessed: 02-May-2023].

[15] R3, "Shaping the Future of Digital Finance & Collaboration | About R3," 2022. [Online]. Available: https://r3.com/company/. [Accessed: 02-May-2023].

[16] NumPy Developers, "numpy.corrcoef," 2022. [Online]. Available: https://numpy.org/doc/stable/reference/generated/numpy.corrcoef.html. [Accessed: 04-May-2023].

[17] Y. Karacaoglu *et al.*, "Blockchain Interoperability (English)," Washington, D.C., 2021.