What Does Matter in the Success of a Decentralized Application? From Idea to Development

Elvira-Maria Arvanitou[1], Dimitrios Gagoutis[2], Apostolos Ampatzoglou[1], Nikolaos Mittas [3], Ignatios Deligiannis[2], Alexander Chatzigeorgiou[1]

[1] Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece
[2] Department of Information & Electronic Engineering, International Hellenic University, Greece
[3] Department of Chemistry, International Hellenic University, Greece

Context: With the rise of blockchain, various applications are running in a decentralized manner, covering the needs of various end-users. Decentralized Applications (DApps) are becoming popular in numerous application domains, ranging from finance to games, and from Non-Fungible Tokens to security mechanisms. The success of a DApp, from a financial perspective, can be perceived as the market fragment that it captures, and the volume of transactions it generates. ***Objective***: The goal of this study is to investigate the factors that are important for safeguarding (as much as possible) the financial success of a Decentralized Application. In this study, we focus on four management factors that could influence financial success: the context of the DApp (e.g., focusing on finance, games, entertainment), the intensity of development activities (e.g., number of: commits, forks, or branches of the repository), the size of the development team and the existence of project documentation. ***Method***: We performed a case study on 122 DApps that were available through an open repository of smart contracts, namely State-of-the-DApps. By mining the repository, we recorded two metrics that capture the financial success of the application (number of users and volume of transactions) and explored their relation to the aforementioned factors. ***Results***: The findings of the study suggest that the intensity of development activities is the most important factor for its financial success. Similarly, the context (i.e., the application domain) of the decentralized application is also a key-factor since it influences the number of users that the DApp will reach. ***Conclusions***: Based on the findings, we suggest businesses that want to enter the market of decentralized applications to balance properly between technical and business parameters. For an application to be successful, it requires both an intensive development process, but also a careful consideration of the application domain.

1. Introduction

Blockchain refers to decentralized and distributed digital systems used to record transactions on a network of computers [8]. A blockchain is essentially a chain of blocks, where each block contains a set of transactions. Once a block is added to the blockchain, it cannot be modified or deleted. This creates a permanent and irrefutable list of all transactions on the blockchain [9]. *Decentralized applications* (DApps) are pieces of software that run on a blockchain. In recent years, DApps have showcased a huge success, with the market being projected to reach 17 billion dollars by the end of 2023 and 35 billion dollars, by the end of 2027 (a CAGR around 18%)[1]. A similar growth can be observed in the interest of the public, as suggested by Google Trends, indicating that the public interest on DApps has increased up to 100 times from 2004 to 2023[2].

[1] https://www.statista.com/outlook/fmo/digital-assets/defi/worldwide
[2] https://trends.google.com/trends/explore?date=all&q=dapps

These latest developments in the landscape of software development open opportunities for businesses, which can significantly benefit by releasing successful DApps, capturing a new market as pioneers. Nevertheless, recent studies (e.g., [6]) suggest that not all decentralized applications are successful, since many of them appear to attract a limited number of users and almost zero transactions. In that sense, businesses must enter this new market carefully, to minimize the development of unsuccessful projects that will result in financial losses and hurt the reputation of the company. To this end, in this study, we investigate various DApps, for which financial indicators are available, and provide an initial exploration of factors that might be related to financial success. As financial success we define the financial benefits (e.g., direct income, promotion to market, etc.) that a DApp can bring to the owner. This benefit is usually proportional to the amount of money that is handled in the DApp transactions. Since this number is not always available for all DApps, we have used two proxies: (a) number of transactions; and (b) number of end-users.

2. Setting the Scene

To investigate the relation among the success of a DApp, the application domain and several aspects associated with the development of the corresponding project, we have relied on a well-known repository of DApps, called the State-of-the-DApps[3]. State-of-the-DApps was initially launched as web portal, from where data could be extracted manually. However, during the data collection phase, the data of the platform was made available as a service, i.e., through API calls, and the web portal has migrated to GitHub. Despite the existence of similar platforms (e.g., DAppRadar), State-of-DApps is the most frequently used in previous research endeavors. State-of-DApps provides generic information on the DApp (e.g., name, application domain), financial information (e.g., number of users and transactions), and a link to the source code repository (e.g., the Git repository of the DApp). The used DApps can be treated as open-source software since their source code is available in a public GitHub repo. Although we cannot claim that this study is the first one that mines DApps repositories for an initial exploration of this field, we believe that this study is novel and emerging in the sense that provides a managerial point of view to the subject, focusing on the financial success of DApps. Wu et al., in two studies [6][7], present a descriptive analysis on the popularity of DApps, summarizing the patterns of using DApps and Smart Contracts to access the underlying blockchain, and explore the worth addressing issues of deploying and operating DApps. Additionally, Cui et al. [1] focus on exploring the activity around a decentralized application, considering as predictors factors such as the number of transactions, development activity, etc. Although this work is close to the goal of our study, Cui et al. use financial indicators as predictors of development intensity, rather than focusing their study on predicting the financial success of the DApp.

3. Methodology

Research Goal and Research Questions: The goal of this study is to study factors that can lead to DApps financial success. As an initial exploration of the problem, in this work we focus on high-level and coarse-grained factors, and we open an interesting line for further research. In this exploratory study, we investigate the relation of four potential success factors (namely: ***Context*** of the ***DApp***, ***Neat Development***, ***Size*** of the ***Development Team***, and ***Existence*** of ***Documentation***) to the actual financial success of existing DApps (captured by the ***Number of Users*** and ***Transactions***). The inclusion of context as a possible

[3] https://state-of-the-dapps.github.io/slate

factor is because different domains might be more fitting than others for launching a DApp (e.g., financial DApps are expected to be more profitable, compared to a DApp for software development); development activity intensity is also considered as a parameter, in the sense that it can act as a proxy of the effort that the team has spent in developing a useful (e.g., many commits are related to the development of richer functionality), as well as the neatness of the development process (e.g., the more the identified issues, the more bugs have been closed before the release). The size of the development team can be a proxy of users, but also an index of the diversity of the people that collaborated in defining the DApp, as well as testing it. Finally, documentation is included, due to a more far-fetched possible relation: a well-documented DApp that clearly describes its operation might be more appealing, in terms of trustworthiness for end-users. To achieve this goal, we performed a case study on real-world DApps, based on the guidelines of Runeson et al. [3]. The protocol and reporting of this case study has been built around a central research question: "*Are the selected factors related to the financial success of a Decentralized Application?"*

Case Selection: The structure of the State-of-DApps at the time of data collection classified applications in 18 categories[4]. To cover all possible application domains, indicating the context of the decentralized application, we recorded data for the 50 most popular DApps from each category. In total, our initial (unfiltered) dataset consisted of 900 applications. The DApps were well-spread across different levels of granularity in terms of size.

Data Collection and Data Pre-Processing: For each DApp (row in our dataset) we recorded various metrics that can be mapped to the tentative success factors and the financial success of the application (columns). For extracting metrics' scores, we have used two sources of information: the State-of-DApps and the Git repository of the corresponding software:

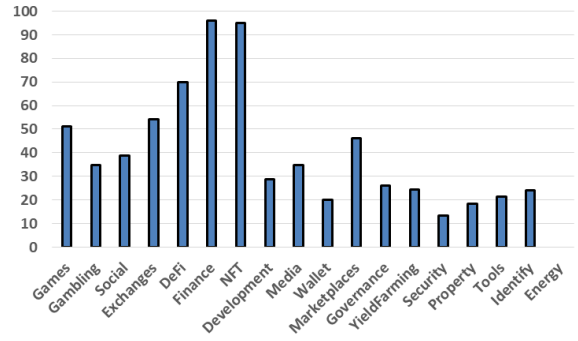
State-of-DApps Metrics	GitHub Metrics
[SD1] Category	[G1] Number of Commits
[SD2] Project Name	[G2] Number of Forks
[SD3] Documentation (Yes / No)	[G3] Number of Issues
[SD4] Number of Users	[G4] Number of Pull Requests
[SD5] Number of Transactions	[G5] Number of Branches
	[G6] Development Team

As a first step towards understanding our sample, in Figure 1, we performed descriptive analytics on the extracted metadata to examine the characteristics of the collected data. As observed from Figures 1.a – 1.c, the distributions of the development activity, transactions, and active users, were quite uneven. The DApps presenting at least one missing value have been removed. After handling the missing values, the dataset comprises of 202 rows and 11 columns. Upon this step, since several categories were represented only by very few instances, we merged the 18 categories into 4 High Level (HL) categories[5], based on the application domain of the project, using the Open Card Sorting method [4]. The final distribution of DApps into HL categories is presented in Figure 1.d. In the final step of pre-processing, to not include

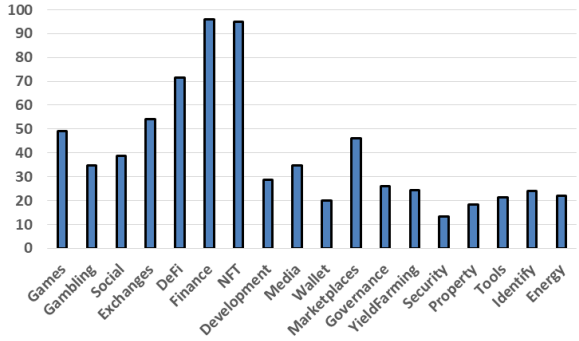
[4] Finance, DeFi, Development, Exchange, Gambling, Game, Governance, Identity, Marketplace, Media, NFT, Property, Security, Social, Wallet, Yield Farming, Tools, and Energy

[5] [Exchange, DeFi, Finance, NFT, Wallet, Energy, Marketplace] → Economy, [Development, Security, Tools, Identity] → Development, [Game, Gambling, Media, Property] → Entertainment, and [Social, Governance] → Social

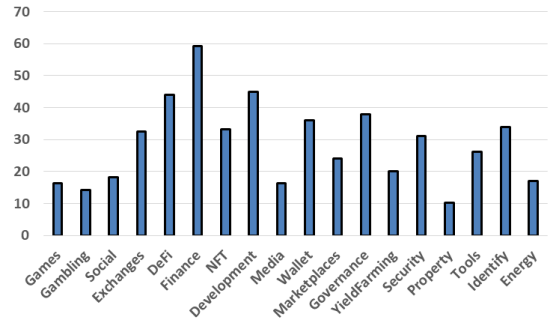
correlated variables, representing the same factor, we merged all GitHub metrics under a cumulative GitHub metric, termed *Code Repository Activity (CRA)* that is the sum of [G1] to [G5].



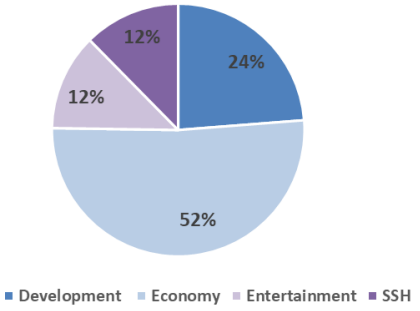
(a) Projects with > 1 transaction



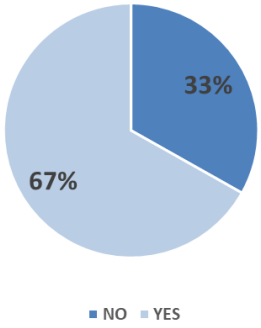
(b) Projects with > 1 active user



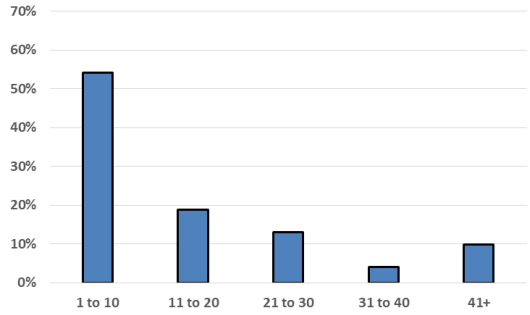
(c) Projects with CRA > 10



(d) Projects per Category



(e) Projects with Available Documentation



(f) Size of Development Team

Figure 1. Sample Analysis

Data Analysis: To investigate the research question that has been set, we have relied on basic descriptive statistics for understanding the characteristics of the collected dataset (e.g., mean scores for the financial success per DApp context, distribution of development events). The model building phase was based on the set of variables presented below:

Dependent Variables (Real-World Financial Success)

[*SD4*] Number of Users: Capturing the Market Segment of the application.

[*SD5*] Number of Transactions: A proxy of the revenue generated by the application.

Independent Variables (Tentative Success Factors)

[*SD1*] Category: Capturing the Idea and the Context of the application.

[*SD3*] Documentation: Indicating the existence of documentation for the application.

[*CRA*] Code Repository Activity: Capturing the intensity of development activities.

[*G6*] Development Team: Capturing the size of the development team.

The inspection of the distribution for both response variables revealed a high degree of skewness, which violates the fundamental assumptions of traditional regression analysis modeling techniques. To alleviate this problem, we decided to transform the dependent variables to 3-point Likert scale ones, based the equal-bins technique [2][6]. After the transformation of both response variables into categorical ones (ordered categories), we performed Ordinal Regression (OR) analysis. Before applying the analysis, we have identified and removed extreme outliers, based on the Tukey's Hinges interquartile range IQR criteria [5][7]. The final dataset, containing information on 122 DApps is available online[8].

4. Results

The fitting of the ordinal regression analysis models revealed a statistically significant improvement against the baseline intercept-only model for both Users ($\chi^2(6) = 44.798, p < 0.001$) and Transactions ($\chi^2(6) = 28.504, p < 0.001$). Concerning the former model (Users), the Wald chi-square test indicated a statistically significant effect of **HL Category** ($\chi^2(3) = 14.558, p = 0.002$), **CRA** ($\chi^2(1) = 24.244, p < 0.001$) and **Documentation** ($\chi^2(1) = 4.353, p = 0.037$) on the number of Users. In contrast, **Team Size** did not present a statistically significant effect on Users $\chi^2(1) = 0.934, p = 0.334$). Interpreting the estimated odds ratio, we can conclude that higher **CRA** is related to greater odds of more Users—see Figure 2.b—($OR \approx 1.0004, 95\%\ CI = [1{,}000, 1.001], p < 0.001$); same as the existence of **Documentation** ($OR = 2.357, 95\%\ CI = [1{,}053, 5.370], p = 0.018$). Finally, *Social* DApps ($OR = 0.096, 95\%\ CI = [0.022, 0.359], p < 0.001$) and *Development* ($OR = 0.320, 95\%\ CI = [0{,}093, 1.041], p = 0.019$) categories decreases the odds for higher number of Users compared to the *Economy* category—Figure 2.a.

As far as the fitting for the Transactions model is concerned, the results revealed a statistically significant effect of **CRA** ($\chi^2(1) = 17.378, p < 0.001$) on the response variable, while the remaining factors seem not to further contribute at an alpha level of 0.05. Beside the overall non-significant effect of **HL Category** ($\chi^2(3) = 5.560, p = 0.135$) on the number of Transactions, we estimated the expected OR for each category against *Economy* (that is the category with the higher frequency of DApps). Indeed, the estimated OR for *Social* ($OR = 0.277, 95\%\ CI = [0.079, 0.892], p = 0.006$) showcased a decrease of the odds for higher values of Transactions compared to *Economy* DApps, a fact that is also portrayed from the analysis of the cross-tabulation matrix (see Figure 2.b).

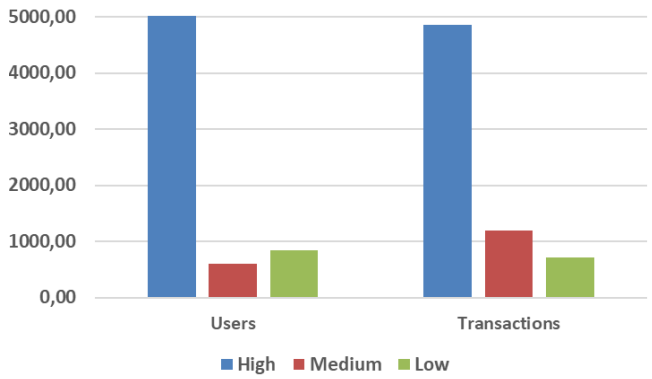
[6] The users' category has been recoded into the following bins: [0, 26], (26, 310], (310, ∞)
The transactions' category has been recoded into the following bins: [0, 75], (75, 1200], (1200, ∞)

[7] value ≥ Q3+ 2×2.1step or value ≤ Q1- 2×2.1step

[8] The dataset is publicly available in https://users.uom.gr/~a.ampatzoglou/aux_material/dappsFinancialFactors.xlsx

Number of Users				Number of Transactions			
Factors	Low	Medium	High	Factors	Low	Medium	High
Development	47%	18%	35%	Development	47%	18%	35%
Economy	28%	35%	38%	Economy	31%	36%	33%
Entertainment	38%	31%	31%	Entertainment	31%	31%	38%
Social	50%	42%	8%	Social	42%	50%	8%

(a) Heat Map on the Relation between Success Factors and HL Categories



(b) Relation between Success Factors and Number of Development Activities

Figure 2. Parameters Interpretation

5. Discussion and Conclusions

This exploratory empirical study, aimed at investigating the relation between the financial success of a DApp and some potential success factors. The analysis was performed on DApps mined through the State-of-DApps repository. The results suggest that ***Code Repository Activity*** and the ***Category*** of the DApp are statistically significant for both models, i.e., related to the number of Users and the number of Transactions. This outcome can be considered intuitive in the sense that these factors capture both the market (**HL Category**) and the technical (**CRA**) perspective. On the one hand, the selection of a fitting target market is important, since it defines both possible sales, as well as the value that a customer is willing to pay. On the other hand, a successful product is expected to have been neatly developed, to provide appropriate features, ease bug fixing before and after release, etc. ***Documentation*** is a statistically significant factor for predicting the number of Users. This can be explained by the fact that DApps are a special case of software. Transparency on how the functionality of a smart contract is implemented, through public documentation, can be important for the adoption of a DApp from end-users. Finally, the ***Size of the Development Team*** does not seem to be significant for predicting the financial success of a DApp.

Threats to Validity: Since this study is an exploratory one, have only studied a limited number of potential factors that affect DApp success, posing a threat to construct validity. The selection of a specific sample before the application of statistical analysis is a threat to validity that applies to any mining repositories empirical study in software engineering. The main mitigation actions that we have taken, for alleviating this threat, are: (a) the specification of concrete project selection criteria; (b) the construction of a large corpus of projects for analysis; and (c) the diversification of the dataset in terms of application domains, size, etc. We believe that given (b) and (c) we have avoided bias as much as possible; whereas

given (a) we report all our decisions so that the reader is informed, and the results are interpretable in the correct context. A final threat to validity is the selection of the DApps repository. We have started our quest for finding a repository of DApps, by searching for such sources in the literature and the web. At the time of the search, we had identified four candidates: the State-of-Dapps, DApps Radar, Web3xplorer, and DApps Review. Given the popularity of the State-of-DApps we have opted towards its use, so that we are certain of the continuity of the existence of our datasource. In addition to that we have observed that State-of-Dapps was the source of information that was the richest at that stage. Currently, out of the examined sources two are unavailable, confirming our choice. Currently, DApps Radar (which was quite a young repository at the point of data extraction) is becoming more mature and deserves future investigation.

Future Work: Future research can focus on the extension of our dataset using more DApps sources (such as DApps Radar, Web3xplorer, etc.) to have a plethora of DApps applications that can be used by researchers and practitioners. Additionally, the investigation of identifying more technical and business factors would be beneficial for both researchers and practitioners. Finally, we plan to perform a qualitative study (i.e., interviews and focus groups) on how the practitioners develop their business strategy, how they define the financial success of their potential apps and what measures they take upfront to guarantee their success.

References

[1] W. Cui, Y. Sun, J. Zhou, Y. Wang, L. Huang, and Y. Yan, "Understanding the Blockchain Eco-system with Analysis of Decentralized Applications: an Empirical Study*", 5th International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS 2021),* 2021.

[2] A. Field, "Discovering Statistics Using IBM SPSS Statistics", SAGE Publications Ltd, November 2017.

[3] P. Runeson, M. Host, A. Rainer, and B. Regnell, "Case study research in software engineering: Guidelines and examples", *Wiley & Sons*, West Sussex, UK, 2012.

[4] D. Spencer, "Card Sorting: Designing Usable Categories", *Rosenfeld Media*, 1st edition, April 2009.

[5] J. W. Tukey, "Exploratory data analysis", *Pearson*, 1st Edition, January 1977.

[6] K. Wu, "An Empirical Study of Blockchain-based Decentralized Applications", *arXiv:1902.04969*, 2019.

[7] K. Wu, Y. Ma, G. Huang, and X. Liu, "A first look at blockchain-based decentralized applications", *Journal of Software: Practice and Experience, Special Issue: Blockchain and Decentralized Applications,* 51 (10), October 2021.

[8] C. Xie, H. Cai, L. Xu, L. Jiang, and F. Bu, "Linked semantic model for information resource toward cloud manufacturing", *IEEE Transactions on Industrial Informatics*, 13(6), pp. 3338–3349, 2017.

[9] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: a survey", *International Journal of Web and Grid Services*, 14(4), October 2018.