Studying How Cryptocurrency Development Characteristics in GitHub Affect Its Market Price and Developer Sentiment in Stack Overflow Discussions

by

Raisul Islam Rashu

A thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of
the requirements for the degree of

Master of Computer Science

in

Computer Science

School of Computer Science
Carleton University
Ottawa, Ontario

January 2020

Copyright ©

2020, Raisul Islam Rashu

Declaration

This is to certify that this thesis is the outcome of the work of the undersigned. I hereby declare that I am the sole author of this thesis. No part of the thesis has been submitted elsewhere, partly or fully for the award of any other degree or diploma. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. Any material reproduced in this thesis has been properly acknowledged.

Raisul Islam Rashu

Student# 101053914

School of Computer Science
Carleton University, Ottawa, Ontario

2020

The undersigned hereby recommend to
the Faculty of Graduate and Postdoctoral Affairs
acceptance of the Thesis

**Studying How Cryptocurrency Development Characteristics
in GitHub Affect Its Market Price and Developer Sentiment
in Stack Overflow Discussions**

Submitted by

Raisul Islam Rashu

in partial fulfilment of the requirements for the degree of
Master of Computer Science

Dr. Olga Baysal, Thesis Supervisor

Dr. Ahmed El-Roby, Internal Examiner

Dr. Mohamed Al Guindy, External Examiner

Dr. Michel Barbeau, Department Chair

Carleton University, 2020

Abstract

Cryptocurrency development has continuous escalation in the past years and holds its presence significantly in open source development. Online collaborative software development platforms such as GitHub offer us an opportunity to observe developer effort, activity and software growth. Cryptocurrency has enabled various applications such as smart contracts, electronically decentralized payments, etc. Since, prices of each cryptocurrency are driven by many factors, we are interested in investigating how various characteristics of cryptocurrency’s codebase development affect market capitalization price. Thus, we conduct a study on a panel dataset containing nearly a year of daily observations of development activity, popularity, and market capitalization for over two hundred open source cryptocurrencies.

Stack Overflow (SO) remains the most popular Q&A forum for software developers, providing solutions to software related problems. SO is a rich source of user specific information with special emphasis on human emotion. In this study, we mine SO data to explore the hot topics related to cryptocurrency development and study the sentiment of cryptocurrency discussions.

Our results demonstrate that 1) the popular cryptocurrencies based on popularity in GitHub are entirely present in the popular cryptocurrency list based on price on CoinMarketCap; 2) **Ethereum** is at the leading position in terms of popularity, while **Bitcoin** dominates in price; 3) using Granger causality analysis, we find no convincing evidence of the “predictive” relationship between software development related metrics such as stars, watchers, forks, contributors, commits, and lines of code changes and market price except for **Ethereum**; 4) developers express a positive sentiment in SO discussions related to popular cryptocurrencies; 5) the extracted keywords from SO discussions related to cryptocurrency prevail the hot topics the developers discuss about.

I dedicate this thesis to my dear parents and elder brother Shariful Islam Raju
who always inspired and believed in me.

Acknowledgments

I, Raisul Islam Rashu, would like to express my deep and sincere gratitude to my supervisor, Professor **Olga Baysal**, for her continuous guidance, patience, insight, helpful discussions throughout my thesis. Her constant encouragement, instructive feedback and support made this work successful. Thanks to the examination committee for their helpful comments.

I am very grateful to my family - my parents and my younger brother Saiful Islam Rimon. They gave me immense support by motivating, energizing and keeping me in their prayers, though being about eight thousands miles apart from me.

I am thankful to my friends and family here in Canada for giving me time, encouragement and support when I needed them most in pursuing my Master's.

Table of Contents

Abstract	iv
Acknowledgments	vi
Table of Contents	vii
List of Tables	xi
List of Figures	xv
List of Acronyms	xix
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Research Questions	4
1.4 Thesis Contributions	7
1.5 Organization of the Thesis	8
2 Related Work	9
2.1 Mining GitHub and Stack Overflow	9
2.1.1 GitHub	9
2.1.2 Stack Overflow	12

2.2	Trends in Cryptocurrency	13
2.3	Social Media Analysis and Cryptocurrency	15
2.4	Sentiment Analysis in Software Engineering	16
3	Methodology	19
3.1	Dataset	19
3.1.1	GitHub Data	20
3.1.2	CoinMarketCap Data	21
3.1.3	Stack Overflow Data	22
3.2	Data Collection and Pre-processing	25
3.2.1	GitHub and CoinMarketCap Data	25
3.2.2	Stack Overflow Data	27
3.3	Data Analysis Techniques	28
3.3.1	Ranking Cryptocurrency Projects	28
3.3.2	Ranking Cryptocurrency Projects' Repositories	30
3.3.3	Stationary and Non-stationary Time Series	32
3.3.3.1	Differencing	34
3.3.3.2	Augmented Dickey-Fuller test	35
3.3.3.3	Kwiatkowski-Phillip-Schmidt-Shin Test	35
3.3.3.4	Types of Stationarity	36
3.3.4	Granger Causality Analysis	37
3.3.5	Sentiment Analysis	38
3.3.6	Word Cloud	39
3.4	Tools Used	40
4	Results	41
4.1	RQ1: What are the current trends in popularity and price of cryptocurrencies?	41

4.2	RQ2: What are the current practices of cryptocurrency development?	51
4.3	RQ3: What are the current trends in development of repositories of popular cryptocurrencies?	53
4.4	RQ4: What are the metrics that may predict cryptocurrency price? .	70
4.5	RQ5: What cryptocurrency dominates SO discussions and what are the hot topics on cryptocurrency that the developers talk about in SO discussions?	77
4.6	RQ6: What is the prevailing sentiment in SO discussions related to cryptocurrency?	80
5	Discussion	84
5.1	Discussion about Findings	84
5.2	Implications for Software Developers	87
5.3	Implications for Researchers	88
5.4	Threats to Validity	90
6	Conclusions	92
6.1	Summary of the Study	92
6.2	Future Work	94
	List of References	95
	Appendix A Ranking Popular Repositories of Popular Cryptocurrencies in GitHub	101
	Appendix B Development Activities of Popular Repositories of Popular Cryptocurrencies in GitHub	117
	Appendix C Word Clouds of Hot topics in SO Discussions Related to Cryptocurrency	168

Appendix D Sentiment Distributions in SO Comments Related to
Cryptocurrency

172

List of Tables

1	List of Acronyms.	xix
2	Metrics collected for each repository every 24 hours and the associated GitHub endpoint	20
3	List of tables in SOTorrent.	23
4	Attributes of the panel dataset.	25
5	Categorized attributes of the panel dataset.	26
6	Attributes of the Tags table.	27
7	Attributes of the Comments table.	27
8	Attributes of the Posts table.	28
9	Attributes of the new Tags table after pre-processing.	29
10	Attributes of the new Posts table after pre-processing.	29
11	Attributes of the new Comments table after pre-processing.	29
12	Final outcome from ADF and KPSS tests.	36
13	Sentiment categories.	39
14	Top 15 cryptocurrency projects based on watchers on 2018-01-21. . .	42
15	Top 15 cryptocurrency projects based on watchers on 2019-02-04. . .	42
16	Top 15 cryptocurrency projects based on stars on 2018-01-21.	43
17	Top 15 cryptocurrency projects based on stars on 2019-02-04.	43
18	Top 15 cryptocurrency projects based on CoinMarketCap rank. . . .	44
19	List of popular common cryptocurrency projects.	45

20	Top 11 popular cryptocurrency projects and number of repositories associated in GitHub.	46
21	Percentage of forked repositories for each of the Top 11 popular cryptocurrency projects in GitHub.	54
22	Popular repositories of Bitcoin based on watchers on 2018-01-21. . .	55
23	Popular repositories of Bitcoin based on watchers on 2019-02-04. . .	55
24	Popular repositories of Bitcoin based on stars on 2018-01-21.	56
25	Popular repositories of Bitcoin based on stars on 2019-02-04.	56
26	Popular repositories of Bitcoin in GitHub.	56
27	List of popular repositories of popular cryptocurrencies in GitHub. . .	63
28	List of popular repositories of popular cryptocurrencies in GitHub. . .	63
29	Results of ADF and KPSS tests of stars for Bitcoin	70
30	Results of ADF and KPSS tests of watchers for Bitcoin	70
31	Results of ADF and KPSS tests of forks for Bitcoin	71
32	Results of ADF and KPSS tests of daily commits for Bitcoin	71
33	Results of ADF and KPSS tests of daily LOCs added for Bitcoin . . .	71
34	Results of ADF and KPSS tests of daily LOCs removed for Bitcoin . .	72
35	Outcomes of Granger causality test for stars.	73
36	Outcomes of Granger causality test for watchers.	74
37	Outcomes of Granger causality test for forks.	74
38	Outcomes of Granger causality test for daily commits.	75
39	Outcomes of Granger causality test for LOCs added.	75
40	Outcomes of Granger causality test for LOCs removed.	76
41	Number of posts related to top 11 popular cryptocurrencies.	79
42	Hot topics in SO discussions.	83
43	Popular repositories of Ethereum based on watchers on 2018-01-21. .	101
44	Popular repositories of Ethereum based on watchers on 2019-02-04. .	101

45	Popular repositories of Ethereum based on stars on 2018-01-21.	102
46	Popular repositories of Ethereum based on stars on 2019-02-04.	102
47	Popular repositories of Ethereum in GitHub.	102
48	Popular repositories of Ripple based on watchers on 2018-01-21.	103
49	Popular repositories of Ripple based on watchers on 2019-02-04.	103
50	Popular repositories of Ripple based on stars on 2018-01-21.	103
51	Popular repositories of Ripple based on stars on 2019-02-04.	103
52	Popular repositories of Ripple in GitHub.	104
53	Popular repositories of Cardano based on watchers on 2018-01-21.	104
54	Popular repositories of Cardano based on watchers on 2019-02-04.	104
55	Popular repositories of Cardano based on stars on 2018-01-21.	105
56	Popular repositories of Cardano based on stars on 2019-02-04.	105
57	Popular repositories of Cardano in GitHub.	105
58	Popular repositories of Litecoin based on watchers on 2018-01-21.	106
59	Popular repositories of Litecoin based on watchers on 2019-02-04.	106
60	Popular repositories of Litecoin based on stars on 2018-01-21.	106
61	Popular repositories of Litecoin based on stars on 2019-02-04.	106
62	Popular repositories of Litecoin in GitHub.	107
63	Popular repositories of EOS based on watchers on 2018-01-21.	107
64	Popular repositories of EOS based on watchers on 2019-02-04.	107
65	Popular repositories of EOS based on stars on 2018-01-21.	108
66	Popular repositories of EOS based on stars on 2019-02-04.	108
67	Popular repositories of EOS in GitHub.	108
68	Popular repositories of Stellar based on watchers on 2018-01-21.	109
69	Popular repositories of Stellar based on watchers on 2019-02-04.	109
70	Popular repositories of Stellar based on stars on 2018-01-21.	109
71	Popular repositories of Stellar based on stars on 2019-02-04.	109

72	Popular repositories of Stellar in GitHub.	110
73	Popular repositories of IOTA based on watchers on 2018-01-21.	110
74	Popular repositories of IOTA based on watchers on 2019-02-04.	110
75	Popular repositories of IOTA based on stars on 2018-01-21.	111
76	Popular repositories of IOTA based on stars on 2019-02-04.	111
77	Popular repositories of IOTA in GitHub.	111
78	Popular repositories of Monero based on watchers on 2018-01-21.	112
79	Popular repositories of Monero based on watchers on 2019-02-04.	112
80	Popular repositories of Monero based on stars on 2018-01-21.	112
81	Popular repositories of Monero based on stars on 2019-02-04.	112
82	Popular repositories of Monero in GitHub.	113
83	Popular repositories of NEO based on watchers on 2018-01-21.	113
84	Popular repositories of NEO based on watchers on 2019-02-04.	113
85	Popular repositories of NEO based on stars on 2018-01-21.	114
86	Popular repositories of NEO based on stars on 2019-02-04.	114
87	Popular repositories of NEO in GitHub.	114
88	Popular repositories of Status based on watchers on 2018-01-21.	115
89	Popular repositories of Status based on watchers on 2019-02-04.	115
90	Popular repositories of Status based on stars on 2018-01-21.	115
91	Popular repositories of Status based on stars on 2019-02-04.	115
92	Popular repositories of Status in GitHub.	116
93	Sentiment Scores.	176

List of Figures

1	The schema of the SOTorrent database	22
2	Venn diagram of selecting popular cryptocurrency projects.	30
3	Venn diagram of selecting popular repositories for popular cryptocur- rency projects.	32
4	Examples of stationary and non-stationary time series.	34
5	Trend in price of top 11 popular cryptocurrencies in the past year. . .	47
6	Trend in price (beyond Bitcoin) of top 11 popular cryptocurrencies in the past year.	47
7	Trend in Price (logged) of top 11 popular cryptocurrencies in the past year.	48
8	Weekly seasonality in price for top 1 to 5 popular cryptocurrencies. .	49
9	Weekly seasonality in price for top 6 to 11 popular cryptocurrencies. .	50
10	Number of GitHub repositories for each programming language for 237 cryptocurrency projects in the past year.	51
11	Number of programming languages used for each top 11 popular cryp- tocurrencies in GitHub in the past year.	52
12	Number of (public) repositories in GitHub for each of top 11 popular cryptocurrencies in the past year.	53
13	Trends in popularity in popular repositories of Bitcoin	58
14	Trends in activity in popular repositories of Bitcoin	59

15	Trends in commits in popular repositories of Bitcoin	60
16	Trends in LOCs added in popular repositories of Bitcoin	61
17	Trends in LOCs removed in popular repositories of Bitcoin	62
18	Sample set of the results of SQL query from Tags table.	76
19	Distribution of SO posts related to top 11 popular cryptocurrencies. .	78
20	SQL query results from Tags table for <i>neo</i> tag.	80
21	Key words in overall SO comments related to Cryptocurrency.	81
22	Distribution of sentiment in SO posts related to top 11 popular cryp- tocourrencies.	82
23	Distribution of sentiment in comments of SO posts related to top 11 popular cryptocurrencies.	82
24	Trends in popularity in popular repositories of Ethereum	118
25	Trends in activity in popular repositories of Ethereum	119
26	Trends in commits in popular repositories of Ethereum	120
27	Trends in LOCs added in popular repositories of Ethereum	121
28	Trends in LOCs removed in popular repositories of Ethereum	122
29	Trends in popularity in popular repositories of Ripple	123
30	Trends in activity in popular repositories of Ripple	124
31	Trends in commits in popular repositories of Ripple	125
32	Trends in LOCs added in popular repositories of Ripple	126
33	Trends in LOCs removed in popular repositories of Ripple	127
34	Trends in popularity in popular repositories of Cardano	128
35	Trends in activity in popular repositories of Cardano	129
36	Trends in commits in popular repositories of Cardano	130
37	Trends in LOCs added in popular repositories of Cardano	131
38	Trends in LOCs removed in popular repositories of Cardano	132
39	Trends in popularity in popular repositories of Litecoin	133

40	Trends in activity in popular repositories of Litecoin	134
41	Trends in commits in popular repositories of Litecoin	135
42	Trends in LOCs added in popular repositories of Litecoin	136
43	Trends in LOCs removed in popular repositories of Litecoin	137
44	Trends in popularity in popular repositories of EOS	138
45	Trends in activity in popular repositories of EOS	139
46	Trends in commits in popular repositories of EOS	140
47	Trends in LOCs added in popular repositories of EOS	141
48	Trends in LOCs removed in popular repositories of EOS	142
49	Trends in popularity in popular repositories of Stellar	143
50	Trends in activity in popular repositories of Stellar	144
51	Trends in commits in popular repositories of Stellar	145
52	Trends in LOCs added in popular repositories of Stellar	146
53	Trends in LOCs removed in popular repositories of Stellar	147
54	Trends in popularity in popular repositories of IOTA	148
55	Trends in activity in popular repositories of IOTA	149
56	Trends in commits in popular repositories of IOTA	150
57	Trends in LOCs added in popular repositories of IOTA	151
58	Trends in LOCs removed in popular repositories of IOTA	152
59	Trends in popularity in popular repositories of Monero	153
60	Trends in activity in popular repositories of Monero	154
61	Trends in commits in popular repositories of Monero	155
62	Trends in LOCs added in popular repositories of Monero	156
63	Trends in LOCs removed in popular repositories of Monero	157
64	Trends in popularity in popular repositories of NEO	158
65	Trends in activity in popular repositories of NEO	159
66	Trends in commits in popular repositories of NEO	160

67	Trends in LOCs added in popular repositories of NEO	161
68	Trends in LOCs removed in popular repositories of NEO	162
69	Trends in popularity in popular repositories of Status	163
70	Trends in activity in popular repositories of Status	164
71	Trends in commits in popular repositories of Status	165
72	Trends in LOCs added in popular repositories of Status	166
73	Trends in LOCs removed in popular repositories of Status	167
74	Key words in comments related to Bitcoin	168
75	Key words in SO comments related to Ethereum	169
76	Key words in SO comments related to Ripple	169
77	Key words in SO comments related to EOS	170
78	Key words in SO comments related to Status	170
79	Key words in SO comments related to Blockchain tag.	171
80	Key words in SO comments related to Cryptocurrency tag.	171
81	Distribution of sentiment in comments of SO posts related to Bitcoin .	172
82	Distribution of sentiment in comments of SO posts related to Ethereum .	173
83	Distribution of sentiment in comments of SO posts related to Ripple .	173
84	Distribution of sentiment in comments of SO posts related to EOS . . .	174
85	Distribution of sentiment in comments of SO posts related to Status .	174
86	Distribution of sentiment in comments of SO posts related to Blockchain tag.	175
87	Distribution of sentiment in comments of SO posts related to Cryptocurrency tag.	175

List of Acronyms

Acronyms	Definition
ADF	Augmented Dickey-Fuller test
API	Application Program Interface
CSV	Comma-Separated Values
GB	Giga Byte
KPSS	Kwiatkowski-Phillips-Schmidt-Shin test
LOC	Lines of Code
MB	Mega Byte
MSR	Mining Software Repository
OSS	Open Source Software
Q&A	Question and Answer
SO	Stack Overflow
UML	Unified Modeling Language
URL	Uniform Resource Locator
XML	Extensible Markup Language

Table 1: List of Acronyms.

Chapter 1

Introduction

In this chapter, we first present our motivation for this thesis (Section 1.1). Section 1.2 outlines our key objectives for the thesis. Section 1.3 states our seven research questions. Section 1.4 offers a list of contributions of this thesis. And finally, Section 1.5 presents the overall structure of the thesis.

1.1 Motivation

Now-a-days people are more comfortable electronically when it comes to monetary transactions. People are using different types of payment such as credit cards, debit cards, email transfers, etc. While everyone prefers to go paperless, *cryptocurrency* [1,2] has become a new medium of transferring funds and making payments or donations. Cryptocurrency, a digital currency (e.g., **Bitcoin**, **Litecoin**, **Ethereum**), is nothing but a project of *blockchain*¹. Blockchain is like a “digital ledger” stored in a distributed network. A block refers to the record of a new transaction. When a transaction (block) is completed, the block is added to the chain. Blockchain is the technology by which cryptocurrency is developed. In the past years, cryptocurrency is developed tremendously. Moreover, cryptocurrency is also used as an investment

¹<https://www.blockchain.com>; accessed on April 3, 2019

similar to a stock market. People buy cryptocurrency with relatively lower rate and sell it when the price increases. So, for a new cryptocurrency developer or an investor, it is necessary to understand the trends in cryptocurrency development. Therefore, we decided to conduct this study to better understand the growth of software development projects related to cryptocurrency and whether their development characteristics affect market price.

*GitHub*² is becoming one of the most popular platforms of hosting software repositories and projects online. The number of repositories hosted in GitHub has now reached over 10 million. In recent years, numerous research studies have been conducted to mine GitHub data to understand trends and development of software systems. Cryptocurrencies hold a significant presence in open source development, comprising some of the most popular repositories on GitHub [1]. Thus, GitHub is a good source of information about the development and evolution of cryptocurrency projects. Cryptocurrency projects on GitHub are our primary targets for exploring cryptocurrency development.

While GitHub provides rich data about development and collaboration on software projects, *Stack Overflow*³ (SO) serves as a venue for developers to post their questions, participate in various discussions, exchange ideas in solving various problems related to software development. SOTorrent⁴ is an open dataset that provides a version history of SO posts, discussions and code snippets. SOTorrent contains various information such as user activities, comments, posts, user views, etc., that could potentially provide user-specific data. SOTorrent is also a rich source to study developer sentiment. *Sentiment analysis* is the study of subjectivity (neutral vs. emotionally loaded) and polarity (positive vs. negative) of a text [3] assigning them with a quantitative mood value [4]. Developer sentiment of SO posts and comments within

²<https://github.com>; accessed on April 3, 2019

³<https://stackoverflow.com>; accessed on April 3, 2019

⁴<https://empirical-software.engineering/projects/sotorrent>; accessed on July 2, 2019

the SO posts could be mined through the sentiment analysis approach to gain insights into human sentiment during project development. Also the hot topics hidden in the SO posts and comments could reveal the recent trends about the cryptocurrency development. Therefore, our research goal also aims at exploring developer sentiment and discussion trends surrounding cryptocurrencies.

1.2 Objectives

Cryptocurrencies and their development have gain everyone's attention in recent years, being featured in the global news, financial sector, and academic research [1]. In the past years, many of the cryptocurrencies embrace open source approach in their development. The presence of cryptocurrency development on GitHub is remarkable. GitHub has recently become a popular online collaborative development platform. Developers use GitHub for storing their projects' code base, as well as for working with the other developers by being able to simultaneously access the projects' version control system. Studies on projects hosted on GitHub are being conducted by researchers because of its popularity, convenience to obtain data, and its distinctive built-in social features. GitHub provides a lightweight mechanism for committing, reviewing and managing code changes. As a result, GitHub has been found to be a transparent development environment with pull request based workflow. GitHub provides numerous benefits to software development teams by offering features that foster team collaboration and distributed development. So, the considerable presence of cryptocurrency development on GitHub offers an exclusive opportunity to study their related developer accomplishment and software growth.

Stack Overflow (SO) is the most preferred Q&A based platform for software developers. SO is widely used by software developers to share issues faced during software development and beyond. The technical platform provides users with solutions for

problems related to various software development aspects and topics such as programming techniques, algorithms, language specific programming issues, UML diagrams, etc. Majority of the posted solutions contain code snippets with necessary text explanations for users' improved understanding [5]. The SO post discussions along with their change history of each posts, individual texts and code blocks are available via a public dataset called SOTorrent [6].

The goal of this thesis is threefold: 1) by mining GitHub cryptocurrency repositories, we want to investigate the development of these repositories, and study various characteristics such as stars, watchers, forks, commits, contributors, lines of code changes, 2) to study how these characteristics affect cryptocurrency price, and 3) to understand developer sentiment (positive, negative, neutral or emotionally loaded) in SO discussions related to cryptocurrency.

In order to gather more insights about cryptocurrency development community, we extract sentiment from the SO discussions. The sentiment allows us to understand what attitude or feelings the developers express in their posts and/or comments. We believe that attaching sentiment to the SO discussions can provide developers with additional key information on what topics are of negative sentiment (e.g., why developers had negative experience) or what features or functionality of the cryptocurrency development contribute to positive emotions.

Therefore, the objective of this thesis is to answer the research questions listed in the next section (Section 1.3).

1.3 Research Questions

In this thesis, we aim at answering the following research questions (RQs):

- **RQ1: What are the current trends in popularity and price of cryptocurrencies?** Bitcoin has gained the most popularity globally in the cryptocurrency market and we want to find out other promising cryptocurrencies beyond Bitcoin that are gaining popularity among investors, as well as among developers. We want to short list some top cryptocurrency projects in terms of popularity metrics in GitHub and correlate them with the CoinMarketCap⁵ ranking. Also we want to study the price trends and volatility.

The result of this RQ will be important to both investors and developers to get the insights into promising cryptocurrencies beyond Bitcoin.

- **RQ2: What are the current practices of cryptocurrency development?**

We study the development of cryptocurrencies in GitHub by looking at the repositories, their size, contributions, and commit history, and determining programming languages that are used in cryptocurrency development.

Developers that are new comers to the area of crypto development can benefit from the results of this RQ. The RQ1 findings can be helpful to suggest them what programming languages are used for crypto development. This information can help them to get familiar with the crypto-specific languages such as C++, JavaScript, Solidity, etc. so they can start contributing to the projects related to crypto infrastructure and development.

- **RQ3: What are the current trends in development of repositories of popular cryptocurrencies?**

There are many projects that are associated with each cryptocurrency in GitHub. We reveal most popular repositories for each popular cryptocurrency and also show how they evolve and grow over the past year.

⁵<https://coinmarketcap.com>; accessed on March 15, 2019

The results of this RQ may be important to cryptocurrency developers to better understand which cryptocurrency is being more or less actively developed in order to decide what cryptocurrency project they can contribute to.

- **RQ4: What are the metrics that may predict cryptocurrency price?**

Investors or researchers are often interested in developing predictive models of the cryptocurrency's price volatility. Since cryptocurrency price is driven by many factors, we want to determine whether our development characteristics of cryptocurrency can be used as predictors of the changes in price, and if so, what characteristics are more powerful predictors.

The results of this RQ can be interesting to investors as predictive models can be useful to predict the future price of cryptocurrency.

- **RQ5: What cryptocurrency dominates SO discussions and what are the hot topics on cryptocurrency that the developers talk about in SO discussions?**

From SO discussions, we mine all the posts that have tags related to cryptocurrency. We show the number of posts related to each cryptocurrency and the name of the cryptocurrency that has the highest volume of posts in SO discussions. Also, we identify various hot topics on cryptocurrency that define the trends that the developers are talking about.

The result of this RQ can help developers to quickly understand the topics that their fellow developers are discussing on online forums. By determining hot topics, developers can understand the key areas for improving their skills and maximizing the development experience.

- **RQ6: What is the prevailing sentiment in SO discussions related to cryptocurrency?** We apply sentiment analysis on SO discussions (both the

posts and the comments) and show the distribution of developers' sentiment polarity (positive vs. negative), as well as the subjectivity (neutral vs. emotionally loaded).

The results of this RQ may be important to developers as sentiment analysis not only allows developers to better plan their future endeavors, it can also help them understand the overall developer experience in the development of cryptocurrency.

1.4 Thesis Contributions

This thesis makes the following contributions:

1. We categorize the popularity and activity metrics based on the development characteristics of projects in GitHub.
2. We provide a list of popular cryptocurrency projects based on their popularity in GitHub and correlate them with the CoinMarketCap ranking.
3. We present the trends of market price of the popular cryptocurrencies for the past year to get insights into how prices change over time throughout the year.
4. We also reveal the price volatility i.e., weekly seasonality of market price of the top popular cryptocurrencies.
5. We perform statistical analysis to find out whether cryptocurrency development characteristics affect its market price.
6. We reveal whether most of the cryptocurrencies are developed using the state-of-the-art programming languages.
7. We further identify the keywords from the SO discussions surrounding the cryptocurrency development.

8. Moreover, we reveal the subjectivity and polarity of developers' sentiment in SO discussions related to cryptocurrency.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 discusses related work. Then, Chapter 3 explains the dataset we used, as well as defines the methodology we followed in our research study. Chapter 4 presents the results of our analysis, while Chapter 5 discusses our findings, challenges, various implications for developers and researchers, as well as discusses threats to validity. Finally, Chapter 6 summarizes our research work and provides several directions for future work.

Chapter 2

Related Work

In this chapter, we discuss relevant research work related to various aspects of the thesis. Section 2.1 focuses on the research related to mining GitHub (section 2.1.1) and Stack Overflow (section 2.1.2). Section 2.2 discusses trends in cryptocurrency research. Related work about social media analysis on cryptocurrency is described in Section 2.3 and finally, Section 2.4 talks about related research on sentiment analysis in software engineering.

2.1 Mining GitHub and Stack Overflow

2.1.1 GitHub

Many studies are conducted to mine GitHub data to understand the trends and developments of software. Researchers mine the information stored in GitHub's event logs to understand how GitHub users employ the platform to collaborate on software development. Kalliamvakou et al. [7] document the results of an empirical study aimed at understanding the characteristics of the repositories in GitHub and how users take advantage of GitHub's main features such as commits, pull requests, and issues. They show that, while GitHub is a rich source of data on software development, mining

GitHub for research purposes should take various potential perils into consideration and also the majority of the projects are personal and inactive, i.e., GitHub is also being used for storing data and as a web hosting service; and that almost 40% of all pull requests do not appear as merged, even though they were.

In another study, Kalliamvakou et al. [8] report findings from an online survey and interviews with GitHub users on how GitHub is used for collaboration in commercial projects and find that many commercial projects adopted practices that are more typical of OSS projects including reduced communication, more independent work, and self-organization. They discuss how GitHub’s transparency and popular workflow can promote open collaboration, allowing organizations to increase code reuse and promote knowledge sharing across their teams.

Georgios and Diomidis [9] present the GHTorrent project, an effort to bring GitHub’s rich product and process data to the hands of the research community, in a scalable manner. GHTorrent is a service that gathers event streams and data from the GitHub hosting site and provides those data back to the community in the form of incremental MongoDB data dumps [9]. In another study, Georgios [10] present the dataset details and construction process of GHTorrent and outline the challenges and research opportunities emerging from it. The GHTorrent project collects data for all public projects available on Github.

Tsay et al. [11] present a study of how developers in open work environments evaluate and discuss pull requests, a primary method of contribution in GitHub, analyzing a sample of extended discussions around pull requests and interviews with GitHub developers. Their study shows that developers raised issues around contributions over both the appropriateness of the problem that the submitter attempted to solve and the correctness of the implemented solution and also the level of a submitter’s prior interaction on a project changed how politely developers discussed the contribution and the nature of proposed alternative solutions.

In another study, Treude and Wagner [12] contribute (i) a broad study of parameters to arrive at good local optima for GitHub and Stack Overflow text corpora, (ii) an a-posteriori characterisation of text corpora related to eight programming languages, and (iii) an analysis of corpus feature importance via per-corpus LDA configuration. They find that (1) popular rules of thumb for topic modelling parameter configuration are not applicable to the corpora used in our experiments, (2) corpora sampled from GitHub and Stack Overflow have different characteristics and require different configurations to achieve good model fit, and (3) they can predict good configurations for unseen corpora reliably. These findings support researchers and practitioners in efficiently determining suitable configurations for topic modelling when analysing textual data contained in software repositories.

Montandon et al. [13] evaluate the performance of unsupervised (based on clustering) and supervised machine learning classifiers (Random Forest and SVM) to identify experts in three popular JavaScript libraries: facebook/react, mongodb/node-mongodb, and socketio/socket.io by collecting 13 features about developers activity on GitHub projects, including commits on source code files that depend on these libraries. They propose a method to identify library experts based on clustering feature data from GitHub; by triangulating the results of this method with information available on LinkedIn¹ profiles and show that it is able to recommend dozens of GitHub users with evidences of being experts in the studied JavaScript libraries.

Manes and Baysal [14] investigate how often GitHub developers re-use code snippets from the SO forum, as well as what concepts they are more likely to reference in their code. Their study show that on average developers make 45 references to SO posts in their projects, with the highest number of references being made within the JavaScript code and 79% of the SO posts with code snippets that are referenced in GitHub code do change over time (at least ones) raising code maintainability and

¹<https://ca.linkedin.com/>; accessed on October 20, 2019

reliability concerns.

2.1.2 Stack Overflow

Identifying user activity and interests from Stack Overflow posts consists of mining the user specific information from the SOTorrent dataset. Multiple research studies have been conducted on user interactions within Stack Overflow posts. For example, Wang et al. [15] identifies the top topics discussed on Stack Overflow question and answers, based on developer interactions using Latent Dirichlet Allocation (LDA).

Diamantopoulos et al. [16] explore the potential of mining Stack Overflow answer edits and propose a similarity scheme for edits to answer posts, using their text and code snippets, and applied clustering to extract useful edit patterns. Their findings indicate that the approach can be effective for identifying commonly applied edits, thus illustrating the transformation path from the initial answer to the optimal solution.

Soni and Nadi [17] explore how comments affect answer updates on Stack Overflow, using the SOTorrent dataset. Their study shows that a large number of answers on Stack Overflow are not updated, even when they receive comments that warrant an update.

Chatterjee et al. [18] describe an exploratory study about the potential usefulness and challenges of mining developer Q&A conversations for supporting software maintenance and evolution tools and they designed the study to investigate the availability of information that has been successfully mined from the other developer communications, particularly Stack Overflow. They analyze the characteristics of chat conversations that might inhibit accurate automated analysis and their results indicate the prevalence of useful information, including API mentions and code snippets with descriptions, and several hurdles that need to be overcome to automate mining that information.

Bosu et al. [19] have studied the user inputs on Stack Overflow to understand the dynamics of reputation building and how it could improve reputation score. They observed that answering quickly, contributing to diverse areas, being available in off peak hours increased the chance of building reputation scores. Their study helped to understand how user interactions could be analyzed in terms of Stack Overflow fields to deduce useful information. We have taken a similar approach to identify user's interest from user posts and comments data of Stack Overflow.

2.2 Trends in Cryptocurrency

In recent years, the development in cryptocurrencies has been significant and it got featured in global news, the financial sector, and academic research. For example, Tonder et al. [1] create a data set that marks the first concentrated effort toward high fidelity panel data of cryptocurrency development for a wide range of metrics. They discuss particularly interesting research opportunities for the combination of data, and release new tooling to enable continuing data collection for future research opportunities as development and application of cryptocurrencies mature. This study helped us to understand the recent trends on cryptocurrency development and their dataset could be analyzed to deduce useful information.

Trockman et al. [2] work on the panel dataset of [1] and find that open source project popularity is associated with higher market capitalization, though development activity and quality assurance practices are insignificant variables in their models. They use Granger causality tests to find a dynamic relation between market capitalization and metrics such as daily stars, forks, watchers, commits, contributors, and lines of code changed [2]. We have taken a similar approach on different attribute to identify whether there is any relation between the development activities and price. We also categorize the popularity and activity metrics based on the

development characteristics of projects in GitHub and provide a list of popular cryptocurrency projects based on their popularity in GitHub and correlate them with the CoinMarketCap ranking. We present the trends of market price of the popular cryptocurrencies for the past year to get insights into how prices change over time throughout the year. We also reveal the price volatility i.e., weekly seasonality of market price of the top popular cryptocurrencies.

In another study, Abraham et al. [20] present a method for predicting changes in **Bitcoin** and **Ethereum** prices utilizing Twitter data and Google Trends data. By analyzing tweets, they show that tweet volume, rather than tweet sentiment (which is invariably overall positive regardless of price direction), is a predictor of price direction and by utilizing a linear model that takes as input tweets and Google Trends data, they are able to accurately predict the direction of price changes.

In a study by Balcilar et al. [21], the authors employ a non-parametric causality-in-quantiles test to analyse the causal relation between trading volume and **Bitcoin** returns and volatility, over the whole of their respective conditional distributions. The causality-in-quantiles test reveals that volume can predict returns — except in **Bitcoin** bear and bull market regimes and the result highlights the importance of modelling non-linearity and accounting for the tail behaviour when analysing causal relationships between **Bitcoin** returns and trading volume, i.e., the volume cannot help predict the volatility of **Bitcoin** returns at any point of the conditional distribution.

Baur et al. [22] explore the question of whether **Bitcoin** is a medium of exchange or an asset and more specifically, what is its current usage and what usage will prevail in the future given its characteristics. They analyse the statistical properties of **Bitcoin** and find that **Bitcoin** is uncorrelated with traditional asset classes such as stocks, bonds and commodities both in normal times and in periods of financial turmoil. They also found that the analysis of transaction data of **Bitcoin** accounts

shows that **Bitcoins** are mainly used as a speculative investment and not as an alternative currency and medium of exchange [22].

Catania et al. [23] study the predictability of cryptocurrencies time series and compare several alternative univariate and multivariate models in point and density forecasting of four of the most capitalized series: **Bitcoin**, **Litecoin**, **Ripple** and **Ethereum**. They apply a set of cryptopredictors and rely on Dynamic Model Averaging to combine a large set of univariate Dynamic Linear Models and several multivariate Vector Autoregressive models with different forms of time variation and find that statistical significant improvements in point forecasting when using combinations of univariate models and in density forecasting when relying on selection of multivariate models.

Zhauniarovich et al. [24] study the use of **Bitcoin** to make donations to open source repositories on GitHub and analyze the amount and volume of donations over time, in addition to its relationship to the age and popularity of a repository. They extract their transactions from **Bitcoins** public blockchain and show a limited adoption of **Bitcoin** as a payment method for receiving donations, with nearly 44 thousand deposits adding up to only 8.3 million dollars in the last 10 years. They also show the weak positive correlation between the amount of donations in dollars and the popularity of a repository, with highest correlation ($r=0.013$) associated with number of forks.

2.3 Social Media Analysis and Cryptocurrency

In recent years, people widely use social networks to communicate and to share their views. **Bitcoin** being a decentralized and peer-to-peer cryptocurrency has attracted a large number of users on the web search and social media [25]. Mittal et al. [25] show the correlation among **Bitcoin** price and Twitter and Google search patterns

using Linear regression, polynomial regression, Recurrent Neural Network, and Long Short Term Memory based analysis that concludes that there is a relevant degree of correlation of Google trends and tweet volume data with the price of `Bitcoin`, and with no significant relation with the sentiments of tweets.

Burnie and Yilmaz [26] develop a new approach to temporalizing word2vec-based topic modelling that determines which topics on social media vary with shifts in the phases of a time series to understand potential interactions, i.e., relevant to the highly volatile `Bitcoin` price with its distinct four phases across 2017–2018. They statistically test which words change in frequency between the different stages and compare four word2vec models to assess their consistency in relating connected words in weighted, undirected graphs. They show for words that fall in frequency when prices shift from rising to falling, all eight topics are identified with the four approaches; for words rising in frequency, three out of the five topics remain constant [26]. Those topics are very intuitive and match the actual events in the news.

2.4 Sentiment Analysis in Software Engineering

A recent trend within software engineering research community has emerged to apply sentiment analysis for enhancing software development, productivity and quality. The studies on sentiment analysis show that developers’ emotions have an impact on their performance and productivity. Sentiments of the text extracted from emails, blogs, and social media posts affect human decision making and behavior [25].

Survey conducted by Wrobel [27] shows that emotions such as anger have a positive impact on developer’s productivity, while negative emotions like disgust, frustration has a negative impact on performance. Further, studies investigate the effect of developers’ emotions in social software engineering by applying sentiment analysis on collaborative environments such as Stack Overflow [28], GitHub [4, 29, 30] and

Jira [31, 32].

Studies, similar to our work, have been conducted on GitHub commits to evaluate the relationship of developer emotions with projects, programming language used for development and team geographical distributions [4]. Pletea et al. [30] observed that security related discussions in GitHub were found to consist more of negative emotions in comparison with the other discussions. Novielli et al. [33] report a benchmark study to assess the performance and reliability of three sentiment analysis tools specifically customized for software engineering.

Haddi et al. [34] explore the role of text pre-processing in sentiment analysis, and report on experimental results demonstrate that with appropriate feature selection and representation, the accuracy of sentiment analysis using support vector machines (SVM) in this area may be significantly improved and the level of accuracy achieved is shown to be comparable to the ones achieved in topic categorisation although sentiment analysis is considered to be a much harder problem in the literature.

Singh and Kumari [35] study the effect of pre-processing on Twitter data for the fortification of sentiment classification especially in terms of slang word. They propose a method of pre-processing that relies on the bindings of slang words on the other co-existing words to check the significance and sentiment translation of the slang word. They use n-gram to find the bindings and conditional random fields to check the significance of slang word and find improvements in accuracy of classification on sentiment classification.

Eom et al. [36] empirically investigate the statistical characteristics and predictability of Bitcoin return and volatility. They state that Bitcoin does not show the dynamic property of volatility persistence, contrary to stylized facts in financial time series and also, the auto regressive model using past volatility does not well work in predicting changes in Bitcoin volatility for future periods. They find that

investor sentiment regarding **Bitcoin** has a significant information value for explaining changes in **Bitcoin** volatility for future periods and these results suggest that **Bitcoin** appears to be an investment asset with high volatility and dependence on investor sentiment rather than a monetary asset [36].

Liu et al. [37] propose a model to analyze sentiment of online stock forum and use the information to predict the stock volatility in the Chinese market. They develop a model to compute the sentimental score of each online post related to a particular stock by generating a sentimental dictionary based on financial terms and such sentimental information is represented by two sentiment indicators, which are fused to market data using the Recurrent Neural Networks (RNNs) for stock volatility prediction. Their empirical study shows that, comparing to using RNN only, the model performs significantly better with sentimental indicators [37].

Chapter 3

Methodology

This chapter is organized according to various modules that comprise our techniques used in this work. We start with describing our dataset in Section 3.1, and then, Section 3.2 discusses how we gathered and filtered our data. Section 3.3 presents the overall view of our main techniques, and finally Section 3.4 explains the tools we employed in our work.

3.1 Dataset

In this research, our dataset contains three types of data: i) cryptocurrency development data from GitHub, ii) financial data of cryptocurrencies from CoinMarketCap¹ and iii) Q&A based discussion data related to cryptocurrency from Stack Overflow (SO). However, (i) and (ii) are merged into one dataset², called *panel dataset*, created by van Tonder et al. [1] for the MSR 2019 Data Showcase³ track. The other dataset we used is extracted from SO (mentioned in (iii)). The following sections (sections 3.1.1, 3.1.2, and 3.1.3) describe the GitHub, CoinMarketCap, and SO dataset respectively that we use in our thesis.

¹<https://coinmarketcap.com>; accessed on March 15, 2019

²<https://github.com/rvantonder/CryptOSS>; accessed on April 10, 2019

³<https://2019.msrconf.org/track/msr-2019-Data-Showcase>; accessed on October 10, 2019

Metric	GitHub Endpoint
# of repository stars	
# of repository forks	/:repo
# of repository subscribers	
# of commits in the last 24 hours	/:repo/commits
# of contributors in the last 7 days	
# of lines of code added & deleted in the last 24 hours	/:repo/commits/:sha
# of contributors during repository lifetime	/:repo/stats/contributors
# of commits in the last year	/:repo/stats/participation
# of lines of code added & deleted in the last year	/:repo/stats/code_frequency

Table 2: Metrics collected for each repository every 24 hours and the associated GitHub endpoint [1].

3.1.1 GitHub Data

Software metrics data from GitHub are collected by the authors of [1] as GitHub is the platform where the vast majority of open source cryptocurrency development takes place. This dataset marks the first concentrated effort toward a panel (i.e., longitudinal) dataset of cryptocurrency development for a wide range of metrics. It also offers a quantitative measure of developer activity for budding open source cryptocurrency development [1]. The data consists of the number of stars, forks, subscribers (watchers), contributors, daily commits, lines of code changes for around 7,000 repositories of 237 cryptocurrency projects on GitHub from January 21, 2018 to February 04, 2019. No data of additional cryptocurrencies added once data collection started (i.e., software metrics for new cryptocurrencies launched mid-year 2018 are excluded in the current data set). The dataset is available on Zenodo⁴ as CSV format to download.

⁴<https://zenodo.org/record/2595588>; accessed on April 20, 2019

Table 2 displays the metrics for each repository, each metric is collected every 24 hours. The metrics give us a view of the popularity and growth of projects in terms of developer activity, repository stars, forks, and subscribers. Last update, number of commits, and lines of code added and deleted in the last 24 hours explain recent active development over time. A single repository may comprise multiple branches and these facilitate various development workflows like Gitflow where certain branches may become stale over time. The data is collected from the default branch of the repository as it is recorded by the `:repo` endpoint on GitHub, and GitHub statistics are also calculated on the repository’s default branch (typically the master branch, but sometimes it deviates to accommodate alternative workflows). The *contributors* endpoint yields a list of contributors over a repository’s lifetime, but unfortunately GitHub limits the contributor list size to 100 contributors where there are more than 100 developers in a repository. The *participation* endpoint renders the weekly aggregate number of commits for the past year and the *code_frequency* endpoint imparts the weekly aggregate of code changes over the repository lifetime. Additionally in the dataset there are two pieces of static metadata available for each repository: (1) whether it is a forked repository and (2) the programming language (determined by GitHubs Linguist library⁵).

3.1.2 CoinMarketCap Data

CoinMarketCap⁶ is currently one of the leading websites for indexing cryptocurrencies. Cryptocurrency market capitalization rank, price, and volume of daily data for each cryptocurrency mentioned in 3.1.1 are collected from CoinMarketCap by the authors of [1]. The cryptocurrency rankings are recorded every 24 hours (downloaded from CoinMarketCap), which incorporate symbol, market capitalization rank, price,

⁵<https://help.github.com/articles/about-repository-languages>; accessed on May 30, 2019

⁶<https://coinmarketcap.com>; accessed on March 15, 2019

Figure 1: The schema of the SOTorrent database [38].

and volume for each cryptocurrency.

3.1.3 Stack Overflow Data

The SOTorrent⁷ dataset is provided by the Mining Software Repository (MSR) Mining Challenge 2019 [39]. The SOTorrent is an open dataset based on the official SO data dump⁸. The MSR 2019 Mining Challenge showcases the SOTorrent dataset for software engineering community to explore. The database contains abundant information about code snippets for bugs, domain specific question and answers, documentation for new API versions and solutions to programming problems. Scenarios

⁷<https://empirical-software.engineering/projects/sotorrent>; accessed on July 2, 2019

⁸<https://archive.org/details/stackexchange>; accessed on July 3, 2019

Tables	
Users	PostVersionUrl
Badges	PostBlockType
Votes	PostHistory
Tags	PostHistoryType
Posts	PostBlockDiff
PostType	PostBlockDiffOperation
Comments	TitleVersion
CommentUrl	PostBlockVersion
PostLinks	GHMatches
PostReferenceGH	Threads
PostVersion	VoteType

Table 3: List of tables in SOTorrent.

like bug fix, documentation update, Q&A clarification make Stack Overflow posts to evolve over time. SOTorrent tracks the history of posts at individual text and code block levels along with user-specific information, as well as provides access to the version history of SO content at the level of whole posts and individual text or code blocks [6]. The dataset connects SO posts to the other platforms by aggregating URLs from text blocks and comments, and by collecting references from GitHub files to SO posts.

The version of the dataset that we used in this study is based on the SO data dump 2019-06-03. The schema of the database [38] is provided in Figure 1. In Figure 1, the tables from the official Stack Overflow data dump are marked gray, the additional tables are marked blue. This schema is based on version 2018-12-09, we do not show all foreign key constraints. The SOTorrent dataset is available for accessing online via Google BigQuery⁹ as Google BigQuery Dataset¹⁰. Moreover, the dataset is also

⁹<https://bigquery.cloud.google.com>; accessed on July 3, 2019

¹⁰https://bigquery.cloud.google.com/dataset/sotorrent-org:2019_06_21; accessed on July 3, 2019

available on Zenodo¹¹ as CSV and XML database dump to download along with the instructions on how to import the dataset. The dataset, being very large (91.4 GB), comprises 22 tables listed in Table 3.

The SOTorrent dataset contains all the tables from the official SO data dump¹² along with few additional tables in the dataset: `PostVersion`, `PostBlockVersion`, `PostBlockDiff`, `TitleVersion`, `PostVersionUrl`, `CommentUrl`, `PostReferenceGH`, `GHMatches`, `PostBlockDiffOperation`, `PostBlockType`, `PostHistoryType`, `PostType`, `VoteType`.

Here, `Post` table consists of the HTML content of the most recent version of SO posts. `PostVersion` table contains version history on the level of whole SO posts. `PostBlockVersion` table is the version history on the level of post blocks, which are either text (1) or code (2) blocks. `PostBlockDiff` table is the line-based difference between post block versions. `TitleVersion` table contains version history of question titles. `PostVersionUrl` table contains all the URLs extracted from text block versions. `CommentUrl` table contains URLs extracted from post comments. `PostReferenceGH` table contains URLs extracted from GitHub repositories pointing to SO posts. `GHMatches` preserves the matched source lines used to build `PostReferenceGH` table. `PostBlockDiffOperation` contains diff operations provided in `PostBlockDiff` table. `PostBlockType` contains available post block types (text or code). `Badges` contains information about badges awarded to SO users. `Comments` contains comments to SO posts. `PostHistory` contains history of SO posts. `PostLinks` contains related SO posts. `Tags` contains question tags. `Users` contains data on SO users. `Votes` contains information on votes for SO posts. `PostHistoryType` contains description of edit events in `PostHistory` table. `PostType` contains information on available types of SO posts. `VoteType` contains

¹¹<https://zenodo.org/record/1434285>; accessed on July 2, 2019

¹²<https://archive.org/details/stackexchange>; accessed on July 3, 2019

data about available types of vote.

Attribute	
date	contributors_7d
cryptocurrency_name	contributors_all_time
repo_name	commits_1y
is_forked	changes_1y_loc_added
language	changes_1y_loc_removed
last_updated	stars
commits_24h	forks
commits_7d	watchers
changes_24h_loc_added	symbol
changes_24h_loc_removed	market_cap_rank
changes_7d_loc_added	price_usd
changes_7d_loc_removed	market_cap_usd

Table 4: Attributes of the panel dataset.

3.2 Data Collection and Pre-processing

In this thesis, we use two datasets: i) the panel dataset and ii) SOTorrent dataset. The following sections (Section 3.2.1 and Section 3.2.2) describe the data collection and pre-processing processes that we apply to both datasets.

3.2.1 GitHub and CoinMarketCap Data

GitHub data and CoinMarketCap data are merged into one dataset called panel dataset. The panel dataset contains data about daily observations of development activity, popularity, and market capitalization for over two hundred open source

cryptocurrencies. The size of the dataset is 353MB¹³ and contains data from January 21, 2018 to February 4, 2019 for 237 cryptocurrency projects. The data contains 3 million entries (one line per entry). Each entry, corresponds to a single repository, comprises 24 comma-separated fields listed in Table 4. Each date has 7,948 rows.

Attribute	
Activity	Popularity
forks	stars
contributors_7d	watchers
commits_24h	
commits_7d	
commits_1y	
changes_24h_loc_added	
changes_7d_loc_added	
changes_1y_loc_added	
changes_24h_loc_removed	
changes_7d_loc_removed	
changes_1y_loc_removed	

Table 5: Categorized attributes of the panel dataset.

In Table 4, the `is_forked` is boolean, `date`, `cryptocurrency_name`, `repo_name`, `symbol`, `last_updated` and `language` are strings, and all the remaining attributes are integers. The value of `contributors_all_time` is limited to 100 since GitHub provides data for up to 100 contributors per repository. The description of attributes is discussed in Section 3.1.1.

One of our goals is to determine whether popular cryptocurrencies in CoinMarket-Cap have similar popularity trends in development. For that, we categorize the data into two categories: 1) popularity metrics and ii) activity metrics. We also discard

¹³The raw compressed data is 5.1GB

the attributes that are not relevant or related to these categories. The categorized metrics that are relevant are reported in Table 5.

Attribute	Type
Id	INTEGER
Tagname	STRING
Count	INTEGER
ExcerptPostId	INTEGER
WikiPostId	INTEGER

Table 6: Attributes of the `Tags` table.

Attribute	Type
Id	INTEGER
PostId	INTEGER
Score	INTEGER
Text	STRING
CreationDate	DATETIME
UserDisplayName	STRING
UserId	INTEGER

Table 7: Attributes of the `Comments` table.

3.2.2 Stack Overflow Data

In order to find answers to our research questions, we require only three tables from the SOTorrent dataset: 1) `Tags`, 2) `Posts`, and 3) `Comments`. Thus, we focused only on these tables since we are interested in SO discussions which can be extracted from these three tables. The attributes of `Tags`, `Posts`, and `Comments` tables are listed in Table 6, Table 8 and Table 7, respectively. SOTorrent contains information about 55,029 tags, 44,945,354 posts, and 73,799,233 comments. The sizes of `Tags`, `Posts`, and `Comments` tables are 2.15MB, 53.7 GB, and 13.1 GB, respectively. Computation

Attribute	Type	Attribute	Type
Id	INTEGER	LastEditorUserId	INTEGER
PostTypeId	INTEGER	LastEditorDisplayName	STRING
AcceptedAnswerId	INTEGER	LastEditDate	DATETIME
ParentId	INTEGER	LastActivityDate	DATETIME
CreationDate	DATETIME	Title	STRING
DeletionDate	DATETIME	Tags	STRING
Score	INTEGER	AnswerCount	INTEGER
ViewCount	INTEGER	CommentCount	INTEGER
Body	STRING	FavoriteCount	INTEGER
OwnerUserId	INTEGER	ClosedDate	DATETIME
OwnerDisplayName	STRING	CommunityOwnedDate	DATETIME

Table 8: Attributes of the **Posts** table.

of such large files is time consuming, thus, we need to minimize the sizes of the tables. So, we eliminate all the irrelevant information from these tables. We use **SQL** queries to make new tables that contain only specific information that is needed to answer our research questions. The attributes of the newly created **Tags**, **Posts**, and **Comments** tables are listed in Table 9, Table 10, and Table 11, respectively.

3.3 Data Analysis Techniques

In the next few sections (Section 3.3.1, Section 3.3.3, Section 3.3.4, Section 3.3.5, and Section 3.3.6), we describe the data analysis techniques that we use in our thesis. In Section 3.4, we discuss the tools we use in our work.

3.3.1 Ranking Cryptocurrency Projects

Now, the data of 237 cryptocurrency projects with 3 million entries (one line per entry or per single repository) can be time consuming and hard to process. Thus, we

Attribute	Type
Id	INTEGER
Tagname	STRING
Count	INTEGER

Table 9: Attributes of the new **Tags** table after pre-processing.

Attribute	Type
Id	INTEGER
Body	STRING
Title	STRING
Tags	STRING

Table 10: Attributes of the new **Posts** table after pre-processing.

Attribute	Type
Id	INTEGER
PostId	INTEGER
Text	STRING

Table 11: Attributes of the new **Comments** table after pre-processing.

focus on popularity of projects (RQ1) and select only top popular projects and their corresponding set of repositories. For this, at first, we sort the data by popularity metrics depicted in Table 5 both on the first day of the data (i.e., 2018-01-21) and on the last day of the data (i.e., 2019-02-04) and list top 15 cryptocurrency by each popularity metrics (*stars* and *watchers*) for each date. Next, we sort the data again by CoinMarketCap Rank (*market_cap_rank*) both on the first day of the data (i.e., 2018-01-21) and on the last day of the data (i.e., 2019-02-04) and list top 15 cryptocurrency for each date. Therefore, we identify popular cryptocurrencies according to three criteria:

1. Popularity (Watchers):

- (a) Top 15 Cryptocurrencies based on *watchers* on 2018-01-21
- (b) Top 15 Cryptocurrencies based on *watchers* on 2019-02-04

2. Popularity (Stars):

- (a) Top 15 Cryptocurrencies based on *stars* on 2018-01-21

(b) Top 15 Cryptocurrencies based on *stars* on 2019-02-04

3. Popularity (CoinMarketCap Rank):

(a) Top 15 Cryptocurrencies based on *market_cap_rank* on 2018-01-21

(b) Top 15 Cryptocurrencies based on *market_cap_rank* on 2019-02-04.

The reason behind this selection process is to determine the common top popular projects that made through the top popularity list throughout the timeline of the dataset. This process can be described by Venn diagram in Figure 2.

Figure 2: Venn diagram of selecting popular cryptocurrency projects.

3.3.2 Ranking Cryptocurrency Projects' Repositories

Each cryptocurrency project comprises multiple repositories on GitHub. Thus, we rank popular repositories from each cryptocurrency project in terms of popularity

metrics (i.e., stars and watchers). For convenience, we limit to top 10 repositories for each cryptocurrency project. Now, at first, we subset the data into each popular cryptocurrency ranked by the method described in Section 3.3.1. Thus, each subset represents all the repositories for that particular cryptocurrency. Then, we sort the data by popularity metrics depicted in Table 5 both on the first day of the data (i.e., 2018-01-21) and on the last day of the data (i.e., 2019-02-04) and list top 10 repositories by each popularity metrics (*stars* and *watchers*) for each date. Thus, we obtain the list of popular repositories of popular cryptocurrencies based on two criteria:

1. Popularity (Watchers):
 - (a) Top 10 repositories of popular cryptocurrencies based on *watchers* on 2018-01-21
 - (b) Top 10 repositories of popular cryptocurrencies based on *watchers* on 2019-02-04
2. Popularity (Stars):
 - (a) Top 10 repositories of popular cryptocurrencies based on *stars* on 2018-01-21
 - (b) Top 10 repositories of popular cryptocurrencies based on *stars* on 2019-02-04

The reason for this filtering is to determine the common top popular repositories for each cryptocurrency project that made through the top popularity list throughout the timeline of the dataset. This process can be described by the Venn diagram shown in Figure 3.

Figure 3: Venn diagram of selecting popular repositories for popular cryptocurrency projects.

3.3.3 Stationary and Non-stationary Time Series

In this section, we state few definitions of time series, stationary series, non-stationary series, etc. that are related to our work. This definitions are adopted from [40].

1. **Time Series:** A time series is a set of observations $\{x_t\}$, each one being recorded at a specific time t [40].
2. **Time Series Model:** A time series model for the observed data $\{x_t\}$ is a specification of the joint distribution (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization. Now we will define the mean function and the covariance function of $\{X_t\}$.

3. **Mean Function of $\{X_t\}$:** Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The mean function of $\{X_t\}$ is

$$\mu_X(t) = E(X_t) \quad (1)$$

4. **Covariance Function of $\{X_t\}$:** Let $\{X_t\}$ be a time series with $E(X_t^2) < \infty$. The covariance function of $\{X_t\}$ is

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))] \quad (2)$$

for all integers r and s .

5. **Stationary Time Series:** Let $\{X_t\}$ be a time series. Then $\{X_t\}$ is stationary if $(X_1, \dots, X_n)$ and $(X_{1+h}, \dots, X_{n+h})$ have the same joint distributions for all integers h and $n > 0$. So we can say that, a series is stationary if its properties (mean, variance and covariance) do not vary with time.
6. **Non-stationary Time series:** Let $\{X_t\}$ be a time series. Then $\{X_t\}$ is non-stationary if $(X_1, \dots, X_n)$ and $(X_{1+h}, \dots, X_{n+h})$ does not have the same joint distributions for all integers h and $n \leq 0$. So, a series is non-stationary where the properties (mean, variance and covariance) vary with time.

Figure 4 can be helpful to understand stationary and non-stationary series. In Figure 4, from series (i) of the first plot, we can clearly see that the mean varies (increases) with time which results in an upward trend. Thus, this is a non-stationary series. In the series (ii) of the first plot, we certainly do not see a trend in the series, but the variance of the series is a function of time. A stationary series must have a constant variance. In the series (iii) of the first plot, the spread becomes closer as the time increases, which implies that the covariance is a function of time. The three examples shown in first plot (a) represent non-stationary time series. Now if we look

at the second plot, in this case, the mean, variance and covariance are constant with time. So the second plot (b) represents stationary time series.

Figure 4: Examples of stationary and non-stationary time series.

Most statistical models require the series to be stationary to make effective and precise predictions. In the next few sections (Section 3.3.3.1, Section 3.3.3.2, and Section 3.3.3.3), we will describe the methods that we follow to determine whether a time series is stationary and also if a series is found to be non-stationary, then how we transform those series into stationary series.

3.3.3.1 Differencing

The idea of transforming a non-stationary time series $\{X_t\}$ into a stationary time series by considering its change series $\{C_t\}$ defined by

$$C_t = X_t - X_{t-1} \quad (3)$$

is called *differencing* [40] where X_t is the value at time t .

3.3.3.2 Augmented Dickey-Fuller test

The Augmented Dickey-Fuller (ADF) test [41] is a statistical test that is used to determine the presence of unit root in the series, and hence help us understand if the series is stationary or not. The null and alternate hypotheses of this test are as follows [42]:

- **Null Hypothesis:** The series has a unit root (value of $\alpha = 1$).
- **Alternate Hypothesis:** The series has no unit root.

If we fail to reject the null hypothesis, we can say that the series is non-stationary [42]. This means that the series can be linear or difference stationary. We discuss more about difference stationary in Section 3.3.3.4.

3.3.3.3 Kwiatkowski-Phillip-Schmidt-Shin Test

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [43] is another test for checking the stationarity of a time series. It is a less popular test than the ADF test. The authors of the KPSS test have defined the null hypothesis as the process is trend stationary, to an alternate hypothesis of a unit root series. The null and alternate hypothesis for the KPSS test are exactly opposite to the ADF test, which often creates confusion. These hypotheses are stated as follows [42]:

- **Null Hypothesis:** The process is trend stationary.
- **Alternate Hypothesis:** The series has a unit root (series is not stationary).

If we fail to reject the null hypothesis, we can say that the series is stationary [42]. We will define the term “stationarity” in Section 3.3.3.4).

3.3.3.4 Types of Stationarity

There are different types of stationarity in the time series and we now describe how to interpret the results of the above tests (ADF and KPSS). The stationarity types are as follows [42]:

1. **Strict Stationary:** For a strict stationary series, its properties (mean, variance and covariance) are not the function of time, that means, a strict stationary series satisfies the mathematical definition of a stationary process.
2. **Trend Stationary:** A trend stationary series is a series that has no unit root but exhibits a trend. If the trend is removed from the series, the resulting series will be strict stationary. The KPSS test classifies a series as stationary on the absence of unit root. This means that the series can be strict stationary or trend stationary.
3. **Difference Stationary:** A time series that can be made strict stationary by differencing is referred to as difference stationary series. ADF test is also known as a difference stationarity test.

Outcomes			
No.	ADF	KPSS	Final
1	Non-stationary	Non-stationary	Non-stationary
2	Stationary	Stationary	Stationary
3	Non-stationary	Stationary	Trend stationary
4	Stationary	Non-stationary	Difference stationary

Table 12: Final outcome from ADF and KPSS tests.

We make sure that the series is truly stationary by applying both ADF and KPSS tests. The possible outcomes of applying the ADF and KPSS stationary tests are presented in Table 12.

From Table 12, we can see that if the outcome of KPSS test is stationary and the outcome of ADF test is not stationary, then the final outcome is trend stationary and by removing the trend it can be transformed into strict stationary series. On the other hand, if the outcome of KPSS test is not stationary and the outcome of ADF test is stationary, then the outcome is difference stationary and by using differencing the series can be made to stationary series. So, in short, the ADF test has an alternate hypothesis of linear or difference stationary, while the KPSS test identifies trend-stationarity in a series.

In this thesis, we follow two statistical tests mentioned in Section 3.3.3.2 and Section 3.3.3.3 to determine whether a series is stationary. For the ADF test, we choose the number of lags using the Akaike Information Criterion (AIC) [44], and for the KPSS test the selected lags type is “tau” and the lags is chosen “short”. we use “*urca*” library in “R” to conduct the ADF and KPSS tests.

3.3.4 Granger Causality Analysis

The Granger causality analysis is a statistical hypothesis test to determine whether a time series is useful to forecast another time series. A series x_t is said to Granger-cause another series y_t if past values of x_t are useful in forecasting y_t , controlling for past values of y_t [45,46]; e.g., if γ is statistically significant in

$$y_t = \delta_0 + \sum_{j=1}^p \alpha_j y_{t-j} + \sum_{j=1}^p \gamma_j x_{t-j} + \epsilon_t \quad (4)$$

Here, ϵ_t is taken to be white-noise series (white Gaussian random vector). p can be equal to infinity but in practice, due to finite length of the available data, p will be finite. α and γ are the matrix for every j . The definition of causality given above implies that x_t Granger causes y_t provided some α and γ are not zero. In summary, Granger causality test is a limited notion of causality where past values of one series

(x_t) are useful for predicting future values of another series (y_t) , after past values of y_t have been controlled for.

In order to apply a Granger causality test, a stationary series is needed. Thus, we check for stationary series first by applying the ADF and KPSS tests. If any series is found to be non-stationary, we make that series stationary by taking the first differences of their logs. Then, we conduct the Granger causality test. In this thesis, we test whether popularity and activity metrics (x_t) Granger cause cryptocurrency price (y_t) which is a standard technique in econometrics. We use “*lmtest*” library in “R” for Granger causality analysis.

3.3.5 Sentiment Analysis

Sentiment analysis, the study of the subjectivity (neutral vs. emotionally loaded) and polarity (positive vs. negative) of a text, has emerged as the main technique for mining effective states from textual developer-generated content [3]. Sentiment analysis, also known as opinion mining, is the field of study that analyzes people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [47]. The studies on sentiment analysis show that developers’ emotions have an impact on their performance and productivity. In this thesis, we analyze and extract the underlying users’ emotions from SO discussions. We convert both the SO comments and SO posts to “UTF-8” encodings for internationalization. Emoticons were also included in the texts to avoid the removal of important information. Our analysis categorized the sentiment into ten categories depicted in Table 13.

We use “*Syuzhet*” library in “R” to analyze the sentiment hidden inside the discussions. We pass the comments (string) and posts (string) to a function provided by “*Syuzhet*” library to calculate the sentiment. The function returns the sentiment scores for each category in numeric format. In Table 13, the **negative** and **positive**

No.	Category
1	anger
2	anticipation
3	disgust
4	fear
5	joy
6	sadness
7	surprise
8	trust
9	negative
10	positive

Table 13: Sentiment categories.

sentiments are the measure of polarity and the rest are the measure of subjectivity.

3.3.6 Word Cloud

Word cloud is a method that allows us to highlight the most frequently used keywords in a paragraph of texts. Word cloud is also known as a text cloud or a tag cloud. It is a visual representation of text data. The most used keywords stand out better in a word cloud, thus, it adds simplicity and clarity. Word clouds are easy to understand and to be shared. Word clouds are visually effective than a table data. In this thesis, we extract hot topics from SO discussions and represent those topics as a word cloud to visualize the main keywords. Thus, we can better understand what topics the developers talk about in SO discussions. Word clouds can help us to understand the trends in active cryptocurrency development. We use “*tm*” library in “R” to clean the text (preprocessing) to generate hot topics from SO discussions. Also, we use both “*wordcloud*” and “*wordcloud2*” libraries in “R” to generate wordcloud of hot topics mined from SO discussions. Initially, for the preprocessing, we convert the SO comments to “UTF-8” encodings for internationalization. Each comment is treated as a document. Then we make all the letters of the documents to lowercase. Further,

we remove all the (i) punctuations, (ii) numbers, (iii) stopwords, (iv) URLs, and (v) white spaces from the documents. Then we converted all the documents to term document matrices and count the frequency of each term in the document. Here, frequency means how many times that term occurs in the document. Later we sort all the terms by frequency in descending order. Then, using the wordcloud libraries, we make the wordcloud of key words extracted from the comments. These are the key words that the developers talk about in SO discussions.

3.4 Tools Used

In this thesis, we use several tools. *SQL* and *Google BigQuery* are used to pre-process the SOTorrent data, as well as to gather data for answering our research questions. “*R*” *programming language* and *RStudio* are used for data analysis and visualization. We use “*prophet*” library in “R” to find out the trend, weekly seasonality and yearly seasonality in price. To check whether a series is stationary, we use “*urca*” library in “R” for ADF and KPSS tests. Furthermore, we use “*lmtest*” library in “R” for Granger causality analysis. Again, we use “*Syuzhet*” library in “R” to analyze the sentiment hidden inside the SO discussions. Also, We use “*tm*” library in “R” to clean the text (preprocessing) to generate hot topics from SO discussions. For generating wordcloud, we use both “*wordcloud*” and “*wordcloud2*” libraries in “R”. Finally, We use “*ggplot2*” library in “R” to generate the graphs.

Chapter 4

Results

In this chapter, we present the results of our study by answering our research questions (Section 4.1, Section 4.2, Section 4.3, Section 4.4, Section 4.5, and Section 4.6 respectively).

4.1 RQ1: What are the current trends in popularity and price of cryptocurrencies?

There are about 237 cryptocurrency projects in our dataset. However, the first part of our first research question aims at determining the most popular cryptocurrencies and their trends. For the convenience, we limit the number of popular cryptocurrency projects to 15. To answer this question, we sort the data by popularity metrics depicted in Table 5 both on the first day of the data (i.e., 2018-01-21) and on the last day of the data (i.e., 2019-02-04) and list top 15 cryptocurrencies by popularity for each date.

So we follow the methodology of ranking of cryptocurrency described in Section 3.3.1. Table 14 and Table 15 report the results based on the “watchers” popularity, i.e., list (1a) and list (1b) from Section 3.3.1, respectively. Moreover, Table 16 and Table 17 show the results for “stars” popularity based on the list (2a) and list (2b)

Popularity (Watchers) on 2018-01-21	
Cryptocurrency	# of Watchers
Ethereum	8,568
Bitcoin	3,302
Lykke	3,249
Ripple	2,329
Stellar	1,966
Factom	1,518
Status	1,501
EOS	1,177
Cardano	1,159
Litecoin	1,159
MaidSafeCoin	1,041
OmiseGO	918
IOTA	889
BitShares	768
Ethereum Classic	729

Table 14: Top 15 cryptocurrency projects based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Cryptocurrency	# of Watchers
Ethereum	11,501
Lykke	4,873
Status	4,226
Bitcoin	4,217
EOS	2,835
Stellar	2,435
Cardano	2,407
Ripple	2,367
Factom	1,848
IOTA	1,533
MaidSafeCoin	1,376
LBRY Credits	1,197
Steem	1,188
Steem Dollars	1,188
Decentraland	1,186

Table 15: Top 15 cryptocurrency projects based on watchers on 2019-02-04.

from Section 3.3.1, respectively. On the other hand, Table 18 presents the results for “coin market cap” popularity based on the list (3a) and list (3b) from Section 3.3.1.

Therefore, we identify 12 cryptocurrencies by mapping Table 14 and Table 15 (as shown in Table 19). Similarly, we determine 15 cryptocurrencies that are present in both Table 16 and Table 17 (as shown in Table 19). Similarly, we find 15 cryptocurrencies that are common between both columns in Table 18 and presented in Table 19. These 15 cryptocurrencies ranked in top 15 throughout the past year in “coin market cap” rank based on price. Table 19 represents the popular cryptocurrencies based on popularity metrics (Watchers and Stars) and “coin market cap rank”. In Table 19, we find that **Ethereum** is at the leading position in terms of watchers and stars, while

Popularity (Stars) on 2018-01-21	
Cryptocurrency	# of Stars
Ethereum	49,658
Bitcoin	27,981
Ripple	5,122
EOS	4,800
LBRY Credits	4,261
IOTA	3,962
Zcash	3,593
Monero	3,225
Litecoin	3,107
Siacoin	3,066
Cardano	3,013
Stellar	2,948
Status	2,805
Lisk	2,371
NEO	2,316

Table 16: Top 15 cryptocurrency projects based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Cryptocurrency	# of Stars
Ethereum	88,968
Bitcoin	40,449
EOS	15,194
Ripple	6,232
Cardano	5,847
IOTA	5,805
Status	5,313
LBRY Credits	5,182
Stellar	5,132
Monero	4,939
Zcash	4,387
Lisk	3,920
Litecoin	3,820
NEO	3,817
Siacoin	3,803

Table 17: Top 15 cryptocurrency projects based on stars on 2019-02-04.

Bitcoin dominates in price. This seems really an interesting finding. Thus, we have a list of total 42 popular cryptocurrencies (12 by *watchers*, 15 by *stars* and 15 by *market_cap_rank* popularity) where 11 of them are common in all the three lists. This process can be described by Venn diagram as shown in Figure 2. The final list of top 11 popular cryptocurrencies is provided in Table 20. We only select those popular cryptocurrency projects listed in Table 20 for further analysis. After selecting top 11 cryptocurrencies, the number of single repositories is reduced from 3,020,240 to 391,020 (391,020 rows) for the entire dataset.

Based on the list of top 11 popular cryptocurrencies, we present the trend in price for these cryptocurrencies in the past year in Figure 5. From Figure 5, it is clear

Rank	2018-01-21	2019-02-04
1	Bitcoin	Bitcoin
2	Ethereum	Ripple
3	Ripple	Ethereum
4	Bitcoin Cash	EOS
5	Cardano	Bitcoin Cash
6	Litecoin	Litecoin
7	NEM	(private)
8	NEO	TRON
9	EOS	Stellar
10	Stellar	(private)
11	IOTA	Cardano
12	Dash	Binance Coin
13	Monero	Monero
14	TRON	IOTA
15	ICON	Dash

Table 18: Top 15 cryptocurrency projects based on CoinMarketCap rank.

that **Bitcoin** dominates the other cryptocurrencies in term of price difference. Since **Bitcoin** is an obvious outlier, we decided to eliminate **Bitcoin** from the data and generate a new graph depicted in Figure 6. But Figure 6 shows that now **Ethereum** is dominating the others followed by **Monero**, **Litecoin**, **NEO**, and **EOS**. Next, we scale the values by taking the log value of the price and generate a new graph illustrated in Figure 7. Figure 7 shows the descending trends in price for top 11 popular cryptocurrencies.

We are then interested in determining the weekly seasonality in price for all the popular cryptocurrencies depicted in Figures 8 and 9. In Figure 8(a), we can see that the price of **Bitcoin** is higher on both Sunday and Monday, and lower on Friday throughout the week. In Figure 8(b), the price of **Ethereum** is higher on Sunday and lower on Friday throughout the week. Figure 8(c) illustrates that, throughout the

No.	Watchers	Stars	CoinMarketCap Rank
1	Ethereum	Ethereum	Bitcoin
2	Lykke	Bitcoin	Ethereum
3	Status	EOS	Ripple
4	Bitcoin	Ripple	EOS
5	EOS	Cardano	Bitcoin Cash
6	Stellar	IOTA	Litecoin
7	Cardano	Status	Cardano
8	Ripple	LBRY Credits	TRON
9	Factom	Stellar	Stellar
10	IOTA	Monero	Monero
11	Litecoin	Zcash	IOTA
12	MaidSafeCoin	Lisk	Dash
13	N/A	Litecoin	Binance Coin
14	N/A	NEO	NEM
15	N/A	Siacoin	NEO

Table 19: List of popular common cryptocurrency projects.

week, the price of **Ripple** is higher on Monday and lower on Friday. In Figure 8(d), the price of **Cardano** is higher on Sunday and lower on Friday. Figure 8(e) depicts that the price of **Litecoin** is higher on Sunday and lower on Friday.

In Figure 9(a), the price of **EOS** is higher on Sunday and lower on Tuesday throughout the week. Figure 9(b) illustrates the price of **Stellar** is higher on Monday and lower on Friday. In Figure 9(c), the price of **IOTA** is higher on Sunday and lower on Friday. Figure 9(d), illustrates that the price of **Monero** is higher on Monday and lower on Friday. Figure 9(e) depicts that the price of **NEO** is higher on Tuesday and lower on Friday. In Figure 9(f), the price of **Status** is higher on Sunday and lower on Friday throughout the week.

RQ1 Findings: The popular cryptocurrencies based on popularity in GitHub are entirely present in the popular cryptocurrency list based on price on CoinMarketCap.

No.	Cryptocurrency	No. of Repositories
1	Bitcoin	4
2	Ethereum	197
3	Ripple	82
4	Cardano	162
5	Litecoin	38
6	EOS	60
7	Stellar	90
8	IOTA	83
9	Monero	19
10	NEO	17
11	Status	277

Table 20: Top 11 popular cryptocurrency projects and number of repositories associated in GitHub.

We observe a *falling wedge*¹ pattern — a downward trend in price for all the popular cryptocurrencies for the past year, while cryptocurrencies have weekly seasonality of higher price either on Monday or Sunday or both and lower on Friday.

¹https://en.wikipedia.org/wiki/Wedge_pattern; accessed on July 2, 2019

Figure 5: Trend in price of top 11 popular cryptocurrencies in the past year.

Figure 6: Trend in price (beyond Bitcoin) of top 11 popular cryptocurrencies in the past year.

Figure 7: Trend in Price (logged) of top 11 popular cryptocurrencies in the past year.

Figure 8: Weekly seasonality in price for top 1 to 5 popular cryptocurrencies.

Figure 9: Weekly seasonality in price for top 6 to 11 popular cryptocurrencies.

4.2 RQ2: What are the current practices of cryptocurrency development?

For 237 cryptocurrency projects, there are about 7,000 GitHub repositories in our dataset. To answer our second research question, we explore the programming languages that are used to develop cryptocurrency projects. Figure 10 represents the list of programming languages and the number of repositories that are being developed in those programming languages for 237 cryptocurrency projects. Figure 10 shows that *Javascript* dominates in cryptocurrency development followed by *Python*, *C#*, *GO*, *C++*, and *Java*. We also notice that there are some repositories in GitHub with unknown development language.

Figure 10: Number of GitHub repositories for each programming language for 237 cryptocurrency projects in the past year.

Moreover, we mine the number of programming languages associated to each

Figure 11: Number of programming languages used for each top 11 popular cryptocurrencies in GitHub in the past year.

popular cryptocurrency. Figure 11 shows the number of programming languages associated to each popular cryptocurrency. From Figure 11, we find that **Status** has the highest number of programming languages (24 languages) followed by **Cardano** (20 languages), **Ethereum** (19 languages), **IOTA** (15 languages), **EOS** (13 languages), **Ripple** (13 languages) and so on. Surprisingly, we find that **Bitcoin**, being dominant in price, is developed by less number of programming languages (4 languages). On the other hand, we also explore the number of repositories associated to each popular cryptocurrency. Figure 12 illustrates the number of repositories associated to each popular cryptocurrency. From Figure 12, we see that **Status** has the highest number of repositories (277 repositories) followed by **Ethereum** (197 repositories), **Cardano** (162 repositories), **IOTA** (83 repositories), **Ripple** (82 repositories) and so on. Surprisingly, again we find that **Bitcoin** has less number of repositories (4 repositories).

RQ2 Findings: Most of the cryptocurrencies are developed using the state-of-the-art programming languages such as *Python*, *C#*, *GO*, *C++*, *Java*, etc. The

Figure 12: Number of (public) repositories in GitHub for each of top 11 popular cryptocurrencies in the past year.

cryptocurrency **Status** has the highest diversity of languages (24), while **Bitcoin** is developed in 4 languages.

4.3 RQ3: What are the current trends in development of repositories of popular cryptocurrencies?

We list top 11 popular cryptocurrencies and number of repositories associated with popular cryptocurrencies. Table 21 shows the percentage of forked repositories on GitHub related to cryptocurrency. Our findings show that there are a lot of forked repositories present in GitHub. 25% repositories are forked for **Bitcoin**, 7.11% for **Ethereum**, 14.63% for **Ripple**, 12.35% for **Cardano**, 57.89% for **Litecoin**, 6.67% for **EOS**, 13.34% for **Stellar**, 4.82% for **IOTA**, 5.26% for **Monero**, 5.88% for **NEO**, 8.67%

for Status.

No.	Cryptocurrency	Repository		
		Total	Forked	% Forked
1	Bitcoin	4	1	25%
2	Ethereum	197	14	7.11%
3	Ripple	82	12	14.63%
4	Cardano	162	20	12.35%
5	Litecoin	38	22	57.89%
6	EOS	60	4	6.67%
7	Stellar	90	12	13.34%
8	IOTA	83	4	4.82%
9	Monero	19	1	5.26%
10	NEO	17	1	5.88%
11	Status	277	24	8.67%

Table 21: Percentage of forked repositories for each of the Top 11 popular cryptocurrency projects in GitHub.

To answer our third research question, we find out the most popular repositories for popular cryptocurrencies and their trends in development. For the convenience, we limit the number of popular repositories to 10. To answer this question, we sort the data by popularity metrics depicted in Table 5 both on the first day of the data (i.e., 2018-01-21) and on the last day of the data (i.e., 2019-02-04) and list top 10 repositories for each popular cryptocurrency by popularity metrics for each date.

Here, we follow the methodology of ranking of repositories of popular cryptocurrencies described in Section 3.3.2. Table 43 and Table 44 in Appendix A depict the extracted information based on the list (1a) and list (1b) from Section 3.3.2, respectively for cryptocurrency **Ethereum**. Moreover, Table 45 and Table 46 in Appendix A depict the extracted information based on the list (2a) and list (2b) from Section 3.3.2, respectively for cryptocurrency **Ethereum**. Therefore, we find the exact match of repositories mentioned in Table 43, Table 44, Table 45, and Table 46 for 8

common popular repositories of cryptocurrency **Ethereum** (as shown in Table 47 in Appendix A).

Similarly, Table 48 and Table 49 in Appendix A depict the extracted information based on the list (1a) and list (1b) from Section 3.3.2, respectively for cryptocurrency **Ripple**. Moreover, Table 50 and Table 51 in Appendix A depict the extracted information based on the list (2a) and list (2b) from Section 3.3.2, respectively for cryptocurrency **Ripple**. Then, we find the exact match of repositories mentioned in Table 48, Table 49, Table 50, and Table 51 for 9 common popular repositories of cryptocurrency **Ripple** (as shown in Table 52 in Appendix A).

We repeat this process for all the top 11 popular cryptocurrencies to find out their popular repositories in GitHub. This process can be described by the Venn diagram as shown in Figure 3. Tables 22 to 26 and Tables 43 to 92 in Appendix A represent all the outcomes of the process of getting popular repositories for top 11 cryptocurrencies. The final list of popular repositories of top 11 popular cryptocurrencies is provided in Table 27 and Table 28. We only select popular repositories of popular cryptocurrency projects listed in Table 27 and Table 28 for further analysis.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
bitcoin	2,785
bips	458
libblkmaker	31
libbase58	28

Table 22: Popular repositories of **Bitcoin** based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
bitcoin	3,522
bips	622
libblkmaker	38
libbase58	35

Table 23: Popular repositories of **Bitcoin** based on watchers on 2019-02-04.

Since we get the list of popular repositories of top 11 popular cryptocurrencies, we show the trend in development of these popular repositories in the past year. Figures 13 to 17 illustrate the trends in popularity and activity of popular repositories

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
bitcoin	26,020
bips	1,875
libbase58	46
libblkmaker	40

Table 24: Popular repositories of *Bitcoin* based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
bitcoin	36,998
bips	3,312
libbase58	74
libblkmaker	65

Table 25: Popular repositories of *Bitcoin* based on stars on 2019-02-04.

Bitcoin	
No.	Repository
1	bitcoin
2	bips
3	libbase58
4	libblkmaker

Table 26: Popular repositories of *Bitcoin* in GitHub.

of *Bitcoin*. In Figure 13, we can see that popularity metrics (watchers and stars) indicate upward trends. In Figure 14, (a) forks indicate an upward trend and (b) weekly contributors for repository *bitcoin* is over 10 throughout the year and the rest are below 5 on average. Figure 15 shows that (a) the average daily commits for repository *bitcoin* is between 4 to 6 and for others it is below 2, (b) the average weekly commits for repository *bitcoin* is around 40 and the rest are below 10, (c) the trend in commits seems slightly increasing for *bitcoin*, slightly decreasing for *bips*, initially increasing and then flat throughout the year for *libbase58*, and *libblkmaker* shows an upward trend initially but a downward trend later in the timeline. In Figure 16, (a) *bitcoin* shows a decreasing trend for daily LOCs added and *bips* initially shows an increasing trend until July 2018 and then it is in a decreasing trend, (b) for *bitcoin*, a decreasing trend till June 2018 and then an increasing one till August 2018 and

again decreasing till October 2018 and finally increasing afterwards, (c) the overall pattern in LOCs added seems to be in uptrend throughout the year for all the popular repositories. In Figure 17, LOCs removed pattern is fairly stable or slightly rising up throughout the year for all the popular repositories.

Figure 24 to Figure 28 in Appendix B illustrate the trends in popularity and activity of popular repositories of **Ethereum**. In Figure 24, popularity metrics (watchers and stars) indicate upward trends with *go-ethereum* being in the lead. In Figure 25, (a) forks indicate an upward trend and (b) weekly contributors for repositories *go-ethereum* and *solidity* are over 5 throughout the year, *EIPs* is around 5, and the rest are below 4, on average. Figure 26 shows that (a) the average number of daily commits for repository *solidity* is above 4, *textitgo-ethereum* is between 2 to 4, and for the others it is below 3, (b) the average number of weekly commits for repository *solidity* is above 30, and the rest are below 25, (c) the trend in commits seems to be increasing for *solidity*, slightly increasing for *go-ethereum*, decreasing for *pyethereum*, and *web3.js* shows an uptrend around middle of January 2019. In Figure 27, (a) *pyethereum*, *go-ethereum*, and *web3.js* show an increasing trend, and the rest show a decreasing trend for daily LOCs added later in the timeline, (b) similarly, *pyethereum*, *go-ethereum*, and *web3.js* show an increasing trend and the rest show decreasing trend for weekly LOCs added later in the timeline, (c) the overall trend in LOCs added seems slightly increasing throughout the year for all the popular repositories but *web3.js* has a sharp jump in increasing trend around middle of January 2019. In Figure 28, LOCs removed seem to have a slightly increasing trend throughout the year for all the popular repositories and similarly, there is a sharp jump to an increasing trend for *web3.js* around middle of January 2019.

(a) Watchers

(b) Stars

Figure 13: Trends in popularity in popular repositories of Bitcoin.

(a) Forks

(b) Weekly Contributors

Figure 14: Trends in activity in popular repositories of Bitcoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 15: Trends in commits in popular repositories of Bitcoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 16: Trends in LOCs added in popular repositories of Bitcoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 17: Trends in LOCs removed in popular repositories of Bitcoin.

No.	Bitcoin	Ethereum	Ripple	Cardano	Litecoin	EOS
1	bitcoin	go-ethereum	rippled	cardano-sl	litecoin	eos
2	bips	wiki	ripple-client	daedalus	litecore-litecoin	Documentation
3	libbase58	web3.js	ripple-lib	Scorex	litecore-lib	eosjs
4	libblkmaker	mist	ripplecharts-frontend	stack2nix	liteaddress.org-old	eosjs-api
5	N/A	solidity	ripple-dev-portal	N/A	lips	eos-token-distribution
6	N/A	EIPs	ripple-client-desktop	N/A	litecore	eosjs-ecc
7	N/A	pyethereum	rippled-historical-database	N/A	litecoin.org-ancient	N/A
8	N/A	ethereumj	ripple-rest	N/A	p2pool-scanner	N/A
9	N/A	N/A	ripple-data-api	N/A	N/A	N/A
10	N/A	N/A	N/A	N/A	N/A	N/A

Table 27: List of popular repositories of popular cryptocurrencies in GitHub.

No.	Stellar	IOTA	Monero	NEO	Status
1	stellar-core	wallet	monero	neo	status-react
2	go	iri	kovri	neo-gui	react-native-desktop
3	js-stellar-sdk	iota.lib.py	monero-gui	docs	status-go
4	stellar-client	android-wallet-app	research-lab	neo-cli	open-bounty
5	stellard	iota.rs	monero-site	examples-csharp	status-network-token
6	horizon	N/A	meta	proposals	contracts
7	docs	N/A	minimero	neo-vm	wiki.status.im
8	stellar-wallet	N/A	kovri-docs	neo-compiler	N/A
9	stellar-protocol	N/A	N/A	neo.org	N/A
10	N/A	N/A	N/A	neo-devpack-dotnet	N/A

Table 28: List of popular repositories of popular cryptocurrencies in GitHub.

Figures 29 to 33 in Appendix B illustrate the trends in popularity and activity of popular repositories of **Ripple**. In Figure 29, we find that popularity metrics (watchers and stars) indicate upward trends. In Figure 30, (a) forks indicate an uptrend and (b) weekly contributors shows a decreasing trend from November 2018 onwards for all repositories. Figure 31 shows that (a) the average number of daily commits for repository *ripple-dev-portal* and *rippled* are above 0.5, and for the others it is below 0.5, (b) the average number of weekly commits for repository *ripple-dev-portal* and *rippled* are above 1, and only *ripple-dev-portal* shows an upward trend after November 2018, and the rest show downward trend, (c) the trend in commits seems to be slightly increasing for *ripple-lib*, slightly decreasing for *rippled-historical-database* and *ripplecharts-frontend*, initially decreasing and then increasing for *rippled*, and *ripple-dev-portal* shows an upward trend initially but slightly downward trend in the end of the timeline. In Figure 32, (a) *ripple-dev-portal* and *ripple-lib* show an increasing trend after September 2018 for daily LOCs added, and *rippled* initially shows a decreasing trends till June 2018, and then it shows an increasing trend till August 2018, (b) *ripple-dev-portal*, *ripple-lib*, and *rippled* are above 3 for the average number of weekly LOCs added, (c) the overall trend in LOCs added seems slightly increasing throughout the year for all the popular repositories. In Figure 33, LOCs removed have a slightly increasing trend throughout the year for all the popular repositories.

Figures 34 to 38 in Appendix B illustrate the trends in popularity and activity of popular repositories of **Cardano**. In Figure 34, the popularity metrics (watchers and stars) indicate upward trends with *cardano-sl* dominating. In Figure 35, (a) forks indicate an upward trend and (b) weekly contributors for repository *cardano-sl* is between 0 to 20 throughout the year, and the rest are below 6, on average. Figure 36 shows that (a) the average number of daily commits for repository *cardano-sl* is between 0 to 16 and for others it is below 8, (b) the average number of weekly commits

for repository *cardano-sl* fluctuates between 20 to 85, and the rest are below 40, (c) the trend in commits seems to be increasing for *daedalus*, decreasing for *cardano-sl*, the rest are following a flat trend throughout the timeline. In Figure 37, (a) *cardano-sl* shows a decreasing trend for daily LOCs added till April 2018 and then increasing till August 2018 and again decreasing on wards, and the rest show increasing trends till June 2018 and then they show a decreasing trend and again increasing from October 2018 and on wards, (b) for *daedalus* and *stack2nix*, are in uptrend till June 2018 and then decreasing till August 2018 and again increasing afterwards, (c) the overall trend in LOCs added seems to be slightly increasing throughout the year for all the popular repositories. In Figure 38, LOCs removed demonstrate a flat trend throughout the year for *Scorex*, and the rest show an increasing trend.

Figure 39 to Figure 43 in Appendix B illustrate the trends in popularity and activity of popular repositories of *Litecoin*. In Figure 39, (a) repository *litecoin* is dominating and is above 6, and the rest are below 5 for watchers, (b) repository *litecoin* is above 7 showing slightly upward trend, while the rest are below 5 for stars. In Figure 40, (a) forks indicate an upward trend and (b) the average number of weekly contributors is under 0.4 for all repositories. Figure 41 shows that (a) the average number of daily commits is under 0.12, (b) the average number of weekly commits fluctuates between 0 to 4, (c) the trend in commits seems to be decreasing for *litecoin*, *litecore*, and *litecore-litecoin*, initially decreasing till April 2018 and then increasing from May 2018 to June 2018 and flat after that throughout the year for *litecore-lib*, and *lips* shows an uptrend after January 2019. In Figure 42, the daily and weekly LOCs added on average show similar trends in (a) and (b), respectively, and in (c) the overall trend in LOCs added seems to be stable throughout the year for all the popular repositories except *lips* that demonstrate an upward trend after January 2019. In Figure 43, LOCs removed metric shows a similar flat trend as LOCs added throughout the year for all the popular repositories.

Figures 44 to 48 in Appendix B illustrate the trends in popularity and activity of popular repositories of EOS. In Figure 44, we can see that popularity metrics (watchers and stars) indicate upward trends for *eos*, *Documentation*, and *eosjs*, while the rest are flat. In Figure 45, (a) forks indicate an upward trend for *eos*, *Documentation*, and *eosjs*, while the rest are flat, and (b) weekly contributors fluctuate between 0 to 35. Figure 46 shows that (a) the average number of daily commits for repository *eos* is over 5 till July 2018 and then comes to under 5 onwards, and the rest are below 5 throughout the year, (b) the average number of weekly commits metric shows a similar trend as the average number of daily commits, (c) the trend in commits seems to be increasing for *eos*, decreasing for *Documentation* and *eos-token-distribution*, initially increasing till September 2018 and then decreasing throughout the year for *eosjs*, *eosjs-api*, and *eosjs-ecc*. In Figure 47, (a) the average number of daily LOCs added is under 8 for all repositories, (b) the average number of weekly LOCs added is under 12 for all repositories, (c) the trend in LOCs added seems to be increasing throughout the year for *eos*, *eosjs*, *eosjs-ecc*, and *eosjs-api*. In Figure 48, (a) the average number of daily LOCs removed metric is under 8 for all repositories, (b) the average number of weekly LOCs removed is under 10 for all repositories, (c) the trend in LOCs removed seems to be increasing throughout the year for *eos*, *eosjs*, *eosjs-ecc*, and *eosjs-api*.

Figures 49 to 53 in Appendix B illustrate the trends in popularity and activity of popular repositories of Stellar. In Figure 49, we can see that *stellar-core* is dominating by popularity metrics (watchers and stars). In Figure 50, (a) forks indicate upward trends and (b) the weekly contributors are under 5, on average. Figure 51 shows that (a) the average number of daily commits is under 1.2, (b) the average number of weekly commits is under 25, (c) the trend in commits seems to be decreasing for *horizon*, *stellar-client*, and *stellar-wallet*, increasing throughout the year for *js-stellar-sdk*, *stellar-protocol* and *docs*, *go*, and *stellar-core* shows an upward trend

initially but a downward trend later in the timeline. In Figure 52, (a) the daily LOCs added, on average, are under 8, (b) the weekly LOCs added, on average, are under 12, (c) the overall trend in LOCs added seems to be flat for *stellar-client*, *horizon*, *js-stellar-sdk* and *stellar-wallet*, but the rest are in uptrend throughout the year. In Figure 53, (a) the daily LOCs removed, on average, are under 6, (b) the weekly LOCs removed, on average, are under 9, (c) the overall trend in LOCs removed seems to be increasing for *stellar-core*, *go*, *stellar-protocol*, while the rest are flat throughout the year.

Figure 54 to Figure 58 in Appendix B illustrate the trends in popularity and activity of popular repositories of IOTA. In Figure 54, we can see that popularity metrics (watchers and stars) indicate upward trends, and repository *wallet* is in the lead. In Figure 55, (a) forks indicate an upward trend and (b) weekly contributors for repository *iri* is over 1 throughout the year, while the rest are below 1, on average. Figure 56 shows that (a) the average number of daily commits is under 2, (b) the average number of weekly commits for repository *iri* is above 4, and for the rest it is below 5, (c) the trend in commits is downward for all repositories except *iri* which is decreasing till October 2018 and then increasing throughout the year. In Figure 57, (a) the average number of daily LOCs added is under 9, (b) the average number of weekly LOCs added is under 1, (c) the overall trend in LOCs added is increasing throughout the year for *iri* and *iota.lib.py*, while the rest are flat. In Figure 58, the overall trend in LOCs removed seems to be increasing for *iri* and *iota.lib.py* and flat for the rest throughout the year.

Figures 59 to 63 in Appendix B illustrate the trends in popularity and activity of popular repositories of Monero. In Figure 59, the popularity metrics (watchers and stars) indicate upward trends, and *monero* dominates the others. In Figure 60, (a) forks indicate an upward trend and (b) weekly contributors are below 8, on average. Figure 61 shows that (a) the average number of daily commits is below 6, (b) the

average number of weekly commits is under 4, (c) the trend in commits seems to be increasing for *monero* and *monero-gui*, while slightly decreasing for the others. In Figure 62, (a) the average number of daily LOCs added is below 9, (b) the average number of weekly LOCs added is also below 9, (c) the overall trend in LOCs added is stable for *mininero* and *meta*, while it is slightly on the rise throughout the year for the other popular repositories. In Figure 63, (a) the average number of daily LOCs removed is below 11, (b) the average number of weekly LOCs removed is also below 10, (c) the overall trend in LOCs removed seems to be stable for *mininero* while slightly increasing throughout the year for the other popular repositories.

Figures 64 to 68 in Appendix B illustrate the trends in popularity and activity of popular repositories of NEO. In Figure 64, (a) the popularity metric “watchers” for *neo* is above 5 and for the rest it is under 5, (b) the popularity metric “stars” indicates upward trends with *neo* dominating. In Figure 65, (a) forks indicate upward trends and (b) weekly contributors are below 7, on average. Figure 66 shows that (a) the average number of daily commits is below 3, (b) the average number of weekly commits is below 17, (c) the number of yearly commits for *docs* is above 200, while for the rest, it is below 200. In Figure 67, (a) the average number of daily LOCs added is under 6, (b) the average number of weekly LOCs added is under 8, (c) the overall trend in LOCs added is slightly increasing throughout the year for all the popular repositories. In Figure 68, (a) the daily LOCs removed are under 12.5, (b) the average number of weekly LOCs removed is under 9, (c) the overall trend in LOCs added is slightly increasing throughout the year for all the popular repositories with the exception of *examples-csharp*.

Figures 69 to 73 in Appendix B illustrate the trends in popularity and activity of popular repositories of Status. In Figure 69, (a) the popularity metric “watchers” for *status-react* is above 5 and under 5 for the rest, (b) the popularity metric “stars”

indicates upward trends except *wiki.status.im* which is slightly downward, and *status-react* is dominating. In Figure 70, (a) the forks for *status-react* are above 6 and for the rest are under 6, (b) weekly contributors *status-react* are above 6, and for the rest they are under 6, on average. Figure 71 shows that (a) the average number of daily commits is below 5, (b) the average number of weekly commits is below 50, (c) the trend in commits is decreasing for *status-network-token*, *wiki.status.im* and *contracts*, increasing for *react-native-desktop* and *status-react*. In Figure 72, (a) the average number of daily LOCs added is under 10, (b) the average number of weekly LOCs added is also under 10, (c) the overall trend in LOCs added is slightly increasing throughout the year for *status-go*, *react-native-desktop*, *status-react*, while for the rest it is flat except *open-bounty* which is slightly increasing till May 2018 and flat afterwards. In Figure 73, (a) the average number of daily LOCs removed is under 11, (b) the average number of weekly LOCs removed is below 9, (c) the overall trend in LOCs removed is slightly increasing throughout the year for *status-go*, *react-native-desktop*, *status-react*, while for the rest it is flat except *open-bounty* which is slightly increasing till June 2018 and flat afterwards.

RQ3 Findings: All the repositories of top 11 popular cryptocurrencies are being actively developed throughout the past year. There are a lot of forked repositories for cryptocurrencies in GitHub, including 25% of repositories are forked for Bitcoin, 7.11% for Ethereum, 14.63% for Ripple, 12.35% for Cardano, 57.89% for Litecoin, 6.67% for EOS, 13.34% for Stellar, 4.82% for IOTA, 5.26% for Monero, 5.88% for NEO, 8.67% for Status.

4.4 RQ4: What are the metrics that may predict cryptocurrency price?

For this research question, we apply a statistical analysis by using a Granger causality test. Predictive or forecasting causality in econometrics can be tested using Granger causality test that measures the ability to predict the future values of a time series using prior values of another time series. In order to apply a Granger causality test, a stationary time series is needed. Thus, we check for stationary series at first by applying the ADF and KPSS tests. We only consider popularity metrics such as stars and watchers, as well as activity metrics such as forks, daily commits, daily LOCs added and daily LOCs removed, since we want to determine whether they may impact daily cryptocurrency price. Table 29 to Table 34 represent the outcomes of the ADF and KPSS tests for **Bitcoin**.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
Stars	bitcoin	8.7	1.27	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	bips	9.7	1.36	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libbase58	4.6	0.77	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libblkmaker	4.1	1.09	-2.5	-1.9	-1.6	0.21	0.14	0.11	No

Table 29: Results of ADF and KPSS tests of stars for **Bitcoin**.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
Watchers	bitcoin	5.6	1.27	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	bips	6.8	1.18	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libbase58	2.7	0.60	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libblkmaker	1.5	0.6	-2.5	-1.9	-1.6	0.21	0.14	0.11	No

Table 30: Results of ADF and KPSS tests of watchers for **Bitcoin**.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
Forks	bitcoin	7.9	1.30	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	bips	9.8	1.22	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libbase58	3.7	1.26	-2.5	-1.9	-1.6	0.21	0.14	0.11	No
	libblkmaker	5.3	1.07	-2.5	-1.9	-1.6	0.21	0.14	0.11	No

Table 31: Results of ADF and KPSS tests of forks for Bitcoin.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
Commits	bitcoin	-6.7	0.04	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	bips	-12.1	0.09	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libbase58	-13.1	0.06	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libblkmaker	-7.5	0.08	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes

Table 32: Results of ADF and KPSS tests of daily commits for Bitcoin.

For the ADF test, if the test statistic is less than the critical value, we can reject the null hypothesis, i.e., the series is stationary. When the test statistic is greater than the critical value, we fail to reject the null hypothesis i.e., the series is not stationary. For the KPSS test, if the test statistic is greater than the critical value, we reject the null hypothesis, i.e., the series is stationary. If the test statistic is less than the critical value, we fail to reject the null hypothesis, i.e., series is stationary.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
LOC Added	bitcoin	-13	0.07	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	bips	-12.8	0.08	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libbase58	-13.1	0.04	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libblkmaker	-13	0.07	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes

Table 33: Results of ADF and KPSS tests of daily LOCs added for Bitcoin.

Metric	Repository	Test Statistic		Critical Value						Stationary?
		ADF	KPSS	ADF			KPSS			
		%	%	1%	5%	10%	1%	5%	10%	
LOC Removed	bitcoin	-12.7	0.09	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	bips	-12.8	0.09	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libbase58	-13.1	0.03	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes
	libblkmaker	-13	0.07	-2.5	-1.9	-1.6	0.21	0.14	0.11	Yes

Table 34: Results of ADF and KPSS tests of daily LOCs removed for **Bitcoin**.

We follow the same procedure for the remaining popular repositories of the popular cryptocurrencies and determine that the series of stars, watchers, contributors, and forks are non-stationary using the ADF and KPSS tests. However, commits and LOCs changes are already found to be stationary. Thus, we conduct the Granger causality test to find out whether popularity and activity metrics Granger-cause cryptocurrency price.

Table 35 to Table 40 represent the outcomes of the Granger causality tests for popular cryptocurrencies. Here, if the p-value is less than 0.05, we reject the null hypothesis, i.e., the metrics do not Granger-cause price. When the p-value is greater than 0.05, we fail to reject the null hypothesis, i.e., the metrics do Granger-cause price. In Table 35, from the result of the granger causality test, we find Granger causality predictive relation between the popularity metric “stars” and the price for **Ethereum**, but for the other popular cyptocurrencies “stars” have no causality effect on price. Similarly, Table 36 and Table 37 illustrate that there is Granger causality between watchers or forks and the price for **Ethereum**, but no causality between these metrics and the price for the other popular cryptocurrencies. From Tables 38, 39, and 40, we can say that activity metrics such as daily commits, daily LOCs changes (added and/or removed) have no Granger causality effect on price for all the popular cryptocurrencies.

RQ4 Findings: The series of stars, watchers, contributors, and forks are found

Metric	Cryptocurrency	p-value	Do stars Granger-cause price?
Stars	Bitcoin	0.2335	No
	Ethereum	3.23e-06	Yes
	Ripple	1	No
	Cardano	0.9997	No
	Litecoin	0.9999	No
	EOS	0.9997	No
	Stellar	0.9996	No
	IOTA	0.9563	No
	Monero	0.9903	No
	NEO	0.9849	No
	Status	0.9908	No

Table 35: Outcomes of Granger causality test for stars.

non-stationary, while commits and LOCs changes are already stationary. We find no convincing evidence of Granger causality between software related metrics such as stars, watchers, forks, contributors, commits, or lines of code changes and price except for **Ethereum**; only stars, watchers, and forks may explain variation in **Ethereum** price.

Metric	Cryptocurrency	p-value	Do watchers Granger-cause price?
Watchers	Bitcoin	0.2165	No
	Ethereum	8.853e-07	Yes
	Ripple	1	No
	Cardano	0.9571	No
	Litecoin	0.9954	No
	EOS	0.9961	No
	Stellar	0.9947	No
	IOTA	0.964	No
	Monero	0.9894	No
	NEO	0.9368	No
	Status	0.2315	No

Table 36: Outcomes of Granger causality test for watchers.

Metric	Cryptocurrency	p-value	Do forks Granger-cause price?
Forks	Bitcoin	0.2418	No
	Ethereum	0.03293	Yes
	Ripple	1	No
	Cardano	0.9919	No
	Litecoin	0.995	No
	EOS	0.9993	No
	Stellar	0.9888	No
	IOTA	0.9622	No
	Monero	0.9964	No
	NEO	0.9252	No
	Status	0.7672	No

Table 37: Outcomes of Granger causality test for forks.

Metric	Cryptocurrency	p-value	Do commits Granger-cause price?
Commits	Bitcoin	0.2091	No
	Ethereum	0.9999	No
	Ripple	0.9934	No
	Cardano	0.9628	No
	Litecoin	0.1245	No
	EOS	0.8518	No
	Stellar	0.9702	No
	IOTA	1	No
	Monero	0.9983	No
	NEO	0.9964	No
	Status	0.897	No

Table 38: Outcomes of Granger causality test for daily commits.

Metric	Cryptocurrency	p-value	Do LOCs added Granger-cause price?
LOCs Added	Bitcoin	0.9778	No
	Ethereum	1	No
	Ripple	0.9939	No
	Cardano	0.9906	No
	Litecoin	1	No
	EOS	0.9999	No
	Stellar	1	No
	IOTA	1	No
	Monero	0.9973	No
	NEO	0.999	No
	Status	0.9874	No

Table 39: Outcomes of Granger causality test for LOCs added.

Metric	Cryptocurrency	p-value	Do LOCs removed Granger-cause price?
LOCs Removed	Bitcoin	0.699	No
	Ethereum	1	No
	Ripple	1	No
	Cardano	0.9979	No
	Litecoin	0.9999	No
	EOS	0.9999	No
	Stellar	1	No
	IOTA	1	No
	Monero	0.9999	No
	NEO	0.9997	No
	Status	0.9362	No

Table 40: Outcomes of Granger causality test for LOCs removed.

Row	Id	TagName	Count
1	1295	accounting	256
2	10100	numericupdown	256
3	10724	change-tracking	256
4	11208	onmouseout	256
5	19404	wshttpbinding	256
6	27039	gdk	256
7	27925	inkcanvas	256
8	28909	quicklook	256
9	33720	git-fetch	256
10	34394	playing-cards	256
11	35071	uimenucontroller	256
12	36524	sybase-iq	256
13	37694	graph-traversal	256
14	39137	php-include	256
15	60710	open-graph-protocol	256
16	62574	jsonstore	256

Figure 18: Sample set of the results of SQL query from Tags table.

4.5 RQ5: What cryptocurrency dominates SO discussions and what are the hot topics on cryptocurrency that the developers talk about in SO discussions?

In the SOTorrent dataset, **Tags** table contains all the tags related to the posts. We extracted all the tags to determine how many tags are related to cryptocurrencies. To extract the necessary data we use the following SQL query(s) on Google's BigQuery:

```
SELECT Id, TagName, Count /*here Count is an attribute*/
FROM Tags
```

We obtain 55,029 data points as the result for our query. This means that in the SOTorrent dataset the total number of tags is 55,029. Sample set of the results are shown in Figure 18. From the results, we obtain all the tags along with the number of times they were mentioned in SO posts. In the other words, this gives us the number of posts related to that tag. Then we search for the tags that are related to cryptocurrencies. We find that the **TagName** of the posts that are related to the discussions on cryptocurrencies are tagged with each cryptocurrency name and additionally with *cryptocurrency* and *blockchain*. Thus, we mine for all the tags and number of posts tagged with those tags related to top 11 popular cryptocurrencies along with both *cryptocurrency* and *blockchain*. We use the following query to obtain the results:

```
SELECT Id, TagName, Count /*here Count is an attribute*/
FROM Tags
WHERE TagName = "blockchain" OR TagName = "cryptocurrency"
OR TagName = "bitcoin" OR TagName = "ethereum"
```

```

OR TagName = "ripple"    OR TagName = "cardano"
OR TagName = "litecoin"  OR TagName = "eos"
OR TagName = "stellar"   OR TagName = "iota"
OR TagName = "monero"    OR TagName = "neo"
OR TagName = "status"

```

The results of the above query are reported in Table 41. In Table 41, *Ethereum* has the maximum number of posts with 2,323 SO posts, while *Bitcoin* stands in the second position with 1,434 SO posts. On the other hand, *blockchain* has the maximum number of posts (2,749 posts). One interesting finding from this query is that we find no post that is related to cryptocurrency *NEO*. We also verify it with another *SQL* query depicted in Figure 20. Overall, we found that SO discussions have around 8,470 posts related to top 11 popular cryptocurrencies. Figure 19 shows the distribution of SO posts that are related to top 11 popular cryptocurrencies.

Figure 19: Distribution of SO posts related to top 11 popular cryptocurrencies.

Id	TagName	Count	Total Percentage
138028	monero	2	0.02 %
112292	ethereum	2,323	27.43 %
132591	stellar	23	0.27 %
130714	litecoin	25	0.30 %
66179	iota	33	0.39 %
130384	eos	55	0.65 %
67868	bitcoin	1,434	16.93 %
6951	status	1,179	13.92 %
65252	ripple	441	5.21 %
104506	blockchain	2,749	32.46 %
131099	cryptocurrency	195	2.30 %
134124	cardano	11	0.13 %
Total = 8,470			

Table 41: Number of posts related to top 11 popular cryptocurrencies.

Moreover, we extract the key words from the SO discussions that are related to cryptocurrency. Figure 21 shows the results of our hot topics extraction from the SO comments related to cryptocurrency. The size of the word indicates its frequency in the documents. Table 41 describes the number of posts associated with each cryptocurrency topic. Similar to sentiment analysis, for individual categories, we set 50 as the threshold for the number of posts to extract hot topics from. We discard all the discussion that have less than 50 posts. As seen in Figure 21, the dominating key words are question, code, transaction, error, problem, blockchain, bitcoin, contract, help, ripple, status, docker, thanks, etc. Figure 74 to Figure 80 in Appendix C illustrate the key words for individual categories related to cryptocurrency. In Figure 74, for **Bitcoin** the dominating key words are question, code, transaction, error, problem, bitcoin, programming, help, API, python, thanks, etc. Table 42 presents few key words extracted from the Figures from 21 to 80. From Table 42 we can conclude that common key words such as question, code, transaction, error, problem,

Row	Id	TagName	Count
Query returned zero records.			

Figure 20: SQL query results from Tags table for *neo* tag.

contract, help, answer, programming, thanks, etc. dominate SO comments related to cryptocurrency.

RQ5 Findings: Ethereum and Bitcoin being the most discussed cryptocurrencies on SO. We also found that some of the hot topics such as *question*, *code*, *transaction*, *error*, *problem*, *contract*, *help*, *answer*, *programming*, *thanks* dominate the comments in SO discussions related to cryptocurrency.

4.6 RQ6: What is the prevailing sentiment in SO discussions related to cryptocurrency?

For our last research question, the results obtained show a similar distribution of sentiment for all the popular cryptocurrencies. Figure 22 and Figure 23 report the results of our sentiment analysis of SO discussions (posts and comments respectively) related to cryptocurrency. In the both cases, the predominating sentiment is positive. Trust sentiment prevails in the second majority of discussions, while the least number of discussions express disgust related sentiment.

Table 41 indicates the number of posts associated with each cryptocurrency topic.

Figure 22: Distribution of sentiment in SO posts related to top 11 popular cryptocurrencies.

Figure 23: Distribution of sentiment in comments of SO posts related to top 11 popular cryptocurrencies.

No.	Overall Comments	Bitcoin	Ethereum	Ripple	EOS	Status	Blockchain tag	Crypto- currency tag
1	question	question	question	question	question	question	question	question
2	code	code	code	code	eos	code	code	code
3	transaction	transaction	contract	ripple	project	answer	transaction	transaction
4	error	error	transaction	answer	error	status	error	problem
5	problem	problem	error	need	api	error	blockchain	answer
6	blockchain	bitcoin	ethereum	issue	nodemodules	app	help	help
7	bitcoin	programming	version	effect	docker	problem	try	function
8	contract	help	node	problem	problem	solution	answer	error
9	ripple	api	solution	solution	machine	user	chaincode	request
10	status	python	function	share	npm	help	version	library
11	help	need	answer	example	bugfix	server	function	information
12	docker	response	try	library	community	php	issue	try
13	solution	documentation	need	api	pushtransaction	data	programming	programming
14	function	example	issue	support	implementation	try	peer	find
15	thanks	thanks	thanks	thanks	thanks	thanks	thanks	thanks

Table 42: Hot topics in SO discussions.

Chapter 5

Discussion

In this chapter, we discuss our study findings in Section 5.1. We also discuss various implications of our study for software developers (Section 5.2) and future researchers (Section 5.3). We describe threats to the validity of this work in Section 5.4.

5.1 Discussion about Findings

The popularity trends are based on the first day of the data (i.e., 2018-01-21) and the last day of the data (i.e., 2019-01-21). In Figures 5, 6, 7, our findings show that **Bitcoin** dominates the other cryptocurrencies in term of price followed by **Ethereum**, **Monero**, **Litecoin**, **NEO**, **EOS** and so on. Figure 7 shows the trends in price for top 11 popular cryptocurrencies. Overall, the price follows a downward trend for all the popular cryptocurrencies. Cryptocurrencies have weekly seasonality of higher price either on Monday or Sunday or both and lower on Friday.

Figure 10 demonstrates the diversity of programming languages and the number of repositories that are being developed in those programming languages for 237 cryptocurrency projects. Here, we set the threshold of number of repositories to 6 because we want to determine the programming languages that are used for the development of more than 5 repositories in GitHub. We also notice that there are

a few repositories in GitHub for cryptocurrency development which are unknown of development language. One possible explanation of this is that the programming language of a repository is determined by GitHub’s Linguist library¹. GitHub uses this library to determine file languages for syntax highlighting and repository statistics. Some files are hard to identify, and sometimes projects contain more library and vendor files than their primary code. That is why we have “Unknown” programming language for a few projects. Surprisingly, we find that **Bitcoin**, being dominant in price, is developed by fewer number of programming languages (i.e., 4 languages). We found that cryptocurrencies are developed using the state-of-the-art development languages.

Table 21 shows the percentage of forked repositories on GitHub related to cryptocurrency. Our findings show that there are a lot of forked repositories present in GitHub. 25% repositories are forked for **Bitcoin**, 7.11% for **Ethereum**, 14.63% for **Ripple**, 12.35% for **Cardano**, 57.89% for **Litecoin**, 6.67% for **EOS**, 13.34% for **Stellar**, 4.82% for **IOTA**, 5.26% for **Monero**, 5.88% for **NEO**, 8.67% for **Status**. We find that the percentage of forked repositories of **Bitcoin** is highest (25%), while **IOTA** has less percentage of forked repositories (4.82%). We also find that all the repositories of the top 11 popular cryptocurrencies are actively developed throughout the past year.

We apply statistical analysis by using Granger causality test to study “predictive causality” between various factors. In order to apply a Granger causality test, a stationary series is needed. Thus, we check for stationary series at first by applying the ADF and KPSS tests. Our experiment shows that the series of stars, watchers, contributors, and forks are non-stationary using the ADF and KPSS tests, thus we took the first differences of their logs to make them stationary. But commits and LOCs

¹<https://help.github.com/articles/about-repository-languages>; accessed on May 30, 2019

changes are already found to be stationary. Using Granger causality analysis, we find no convincing evidence of the dynamic relation between software related metrics such as stars, watchers, forks, contributors, commits, and lines of code changes and price except for **Ethereum**; only stars, watchers, and forks have causality effect on **Ethereum**'s price.

Developers discuss cryptocurrency topics on SO and **Ethereum** has the maximum number of posts with 2,323 SO posts, while **Bitcoin** follows next having 1,434 SO posts. If we consider all the tags related to cryptocurrencies, **Blockchain** has the maximum number of posts with 2,749 posts. One interesting finding is that we do not find any post related to cryptocurrency **NEO**. Perhaps, **NEO** is not being tagged in the posts, or it is just not being discussed in SO. Most of the cryptocurrencies are being discussed by investors, traders, etc. on more specialized online forums dedicated to cryptocurrencies and blockchain technologies, e.g., *Bitcoin Forum*².

We extract the keywords from the SO discussions for each of top 11 popular cryptocurrencies. Similar to sentiment analysis, for individual categories, we set 50 as a threshold for the number of posts to extract hot topics from. We discard all the discussions that have less than 50 posts. Some common keywords such as *question*, *code*, *transaction*, *error*, *problem*, *contract*, *help*, *answer*, *programming*, *thanks* dominate SO comments related to cryptocurrency. Thus, this shows that developers ask questions regarding their problems. They also talk about errors in their code. They ask help when they account code implementation problems or errors in the code. They also talk about transactions and contracts, as well as thank their fellow developers for the provided help. It seems that after removing the stop words from the SO discussions, we still see some common words such as *question*, *answer*, *can*, *get*, *use*, *just*, *send*, etc. We believe that it would be better to apply a cryptocurrency domain specific list of stop words to remove them from the SO discussions. This would allow us to

²<https://bitcointalk.org/>; accessed on October 20, 2019

have more relevant hot topics that are only related to cryptocurrency development.

We observe that all the popular cryptocurrencies have a similar distribution of sentiment across SO. While the SO discussions are analyzed separately for each topic, the social sentiment of the SO discussions is of similar distribution. For individual topic-wise sentiment analysis, we set 50 as threshold for number of posts to analyze sentiment so that we have enough information for the analysis. We discard all the discussions that have less than 50 posts. Our findings suggest that the cryptocurrency developers' emotions extracted from their SO discussions are overall *positive*. *Trust* sentiment prevails in the second majority of the discussions, while the least number of discussions are related to *disgust* sentiment.

5.2 Implications for Software Developers

Software developers interact with numerous software artifacts and participate in various online communities during their daily development activities. GitHub is a popular social collaborative project development platform. Cryptocurrency holds a significant presence in open source development in GitHub. Our experiment provides information on programming languages by which cryptocurrency projects are being developed. Developers can gain insights about popular programming languages for developing cryptocurrencies. If new developers or experts of the other programming languages are interested in contributing to the development of cryptocurrencies, our study can be helpful to suggest them what programming languages they can start with. Also our study provides information on the forked repositories of popular cryptocurrencies. As a result, developers can better understand which cryptocurrency is being more or less actively developed in order to decide what project they can contribute to.

Our study reveals which cryptocurrencies are actively discussed by developers in

SO. That means that there is more interest in the development of certain cryptocurrencies. The presented technique in our study provides a better way for the developers to get an understanding of the hot topics that their fellow developers are talking about while they are involved in crypto development. This can help developers to quickly understand the topics that the other developers are discussing.

There are various other implications for developers from our study. We have found that sentiment analysis not only allows developers to better plan their future endeavors, it can also help them understand the overall developer experience in the development of cryptocurrency. Our analysis infers that the cryptocurrency developers' emotions are mainly *positive* in the SO discussions. Thus, it seems that developers are generally feeling positive about the development of cryptocurrencies.

5.3 Implications for Researchers

Several previous research studies have been conducted on GitHub data, SO data and sentiment analysis. Major research has been recently focused on GitHub [7, 8, 11–14]. The major focus of these studies has been towards mining GitHub data to understand the trends and developments of software. However, only a few studies have focused on understanding how GitHub users employ the platform to collaborate on software development [8, 11, 14]. On the other hand, major research has been also focused on SO [15–19]. The major focus of these studies has been towards automating and improving the process of development discussions in some way. Several research studies have employed sentiment analysis [4, 25, 27, 30, 33–37]. The main focus of these studies has been towards sentiment analysis for enhancing software development, productivity and quality. Further, studies have investigated the effect of developers' emotions in social software engineering by applying sentiment analysis on collaborative environments such as Stack Overflow [28], GitHub [4, 29, 30] and Jira [31, 32]. We believe

that there is much room in finding better ways of automation and extraction of useful information from the GitHub data, SO data and sentiment analysis.

There have been numerous developments in the field of cryptocurrency development due to the latest trend of people being more and more interested in investing in it. Several previous research studies were conducted on cryptocurrency. Trockman et al. [2] studied the panel dataset of [1] and found that open source project popularity is associated with higher market capitalization, though development activity and quality assurance practices are insignificant variables in their models. In our work, we identify whether there is any causality relation between the development activities and price, as well as popularity and price of cryptocurrencies.

We observe that many cryptocurrencies experienced an overall downward trend in the past year, which has no relation to development characteristics. Thus, academic researchers may continue this line of research and develop new techniques to study relationship between crypto's price and its development characteristics. We understand that the timeline of the panel dataset may not sufficient for studying and understanding the long-term relationship between software quality and financial indicators. Thus, we believe there are many opportunities for future research on cryptocurrency development (e.g., predictive analytics).

While we understand that studying price volatility is very time consuming and require additional expertise in conducting studies, we would like to encourage academic research community to be more engaged with the emerging blockchain technologies and conduct more studies investigating cryptocurrency price volatility, relationship between development metrics and financial indicators, sentiment analysis of various social platforms (e.g., Twitter, Facebook), as well as online forums related to cryptocurrencies (e.g., Bitcointalk.org).

5.4 Threats to Validity

Threats to validity are related to the quality of study design and rigorousness of its execution. In our study, we perform various types of the data analysis. There are several threats to validity for our study. The key threats are related to the data collection, as well as data analysis. In this section, we highlight the conditions under which our study is valid and significant for the research community.

The value of “`contributors_all_time`” in the panel dataset is limited to 100. GitHub limits the contributor list size to 100 contributors per repository. As a result, it was not possible for us to determine the actual number of active developers for a repository. Also, for some days (around 19 days), the daily observations were missed due to the authors’ server outage while collecting the panel dataset. Where possible, the authors of [1] recovered the data from GitHub, GHTorrent, and CoinMarketCap. However, missing data had a negligible impact on our results.

The programming languages is extracted in the panel dataset are determined by GitHubs Linguist library³. Some files are hard to identify, and sometimes projects contain more library and vendor files than their primary code. Therefore, the library can not determine the language of the code base, hence, gives *unknown* programming language. The timeline of the panel dataset may not good for understanding the long-term relationship between software quality and price.

The results that we obtain from our study using the SOTorrent dataset are based on the official Stack Overflow data dump released on 2019-06-03. The updated SOTorrent released on 2019-06-21 is used for the study. The popular cryptocurrency topics deduced from the study are extracted for that specific time frame. The results may vary when a new version of the SOTorrent dataset may be considered for the analysis. Also we collect the data based on `tags` available in SO.

³<https://help.github.com/articles/about-repository-languages/>; accessed on April 3, 2019

There may be many posts that lack tags. Thus, we were not able to fetch those SO posts that had no tag. Moreover, “*Status*” can be ambiguous as a tag in SO discussions since “status” may have different meanings in different perspectives such as git status, status of pull request, status of updates, etc. Therefore, the tag “*Status*” in SO discussions might lead to something else than the name of the cryptocurrency **Status**. For sentiment analysis, stemming is not carried out on comments to avoid loss of useful information. Special characters like emojis are also kept in the comments since we believe that emojis are powerful expressions of emotions.

In our study we only consider SO discussions that relate to cryptocurrency topics. Discussions on SO are typically limited to software development, thus, making it a threat to validity. Some cryptocurrencies and their specific aspects related to trading, investment, financial trends, etc., can be discussed elsewhere, e.g., **Bitcoin** Forum, **Bitcointalk** or the other dedicated forums. In our study, we use the Granger causality test as a proxy for studying causality between two variables in a time series. Causality is closely related to the idea of cause-and-effect, although it is not exactly the same. By our definition, causality is observed if a particular variable comes before another in the time series. In economics, the Granger test is often used to find “predictive causality” or predict the future values of time series using prior values of another time series⁴.

⁴https://en.wikipedia.org/wiki/Granger_causality; accessed on October 20, 2019

Chapter 6

Conclusions

Our study aims at applying mining techniques and data analysis to both the GitHub panel dataset and SOTorrent dataset to study cryptocurrency development characteristics in GitHub and whether these characteristics affect cryptocurrency's market price. We also study developer sentiment in SO discussions related to cryptocurrency and blockchain technologies. Section 6.1 presents a summary of our study and the achieved results, while Section 6.2 outlines possible future directions.

6.1 Summary of the Study

The main goal of this thesis is threefold: 1) by mining GitHub cryptocurrency repositories, we aim to investigate the development of these repositories, and study various characteristics such as stars, watchers, forks, commits, contributors, lines of code changes, 2) to study how these characteristics affect cryptocurrency price, and 3) to understand developer sentiment (positive, negative, neutral or emotionally loaded) in SO discussions related to cryptocurrency.

To achieve this, we conduct a study on a GitHub's panel dataset containing nearly a year of daily observations of development activity, popularity, and market capitalization for over two hundred open source cryptocurrencies. Also in this study, we mine

SO data to explore the hot topics related to cryptocurrency development and study the sentiment of cryptocurrency discussions. We find that the popular cryptocurrencies based on popularity on GitHub are entirely present in the popular cryptocurrency list based on price from the CoinMarketCap. We also find that the price is downtrend for all the popular cryptocurrencies. The price have weekly seasonality of higher price either on Monday or Sunday or both and lower on Friday. We further find that all the repositories of top 11 popular cryptocurrencies are actively developed throughout the past year and they are developed by the state-of-the-art developing languages. We find no convincing evidence of the causality relationship between software related metrics such as stars, watchers, forks, contributors, commits, lines of code changes and price except for **Ethereum**; only stars, watchers, and forks have causality effect on **Ethereum**'s price.

We find that developers discuss cryptocurrency topics on SO and **Ethereum** has the maximum number of posts, while **Bitcoin** stands in the second position. When considering all cryptocurrency and blockchain related tags, **Blockchain** has the highest number of posts. Also, we extract the hot topics from the SO discussions for each of top 11 popular cryptocurrencies. The generated hot topics along with the sentiment can help developers to quickly understand key points discussed by the other developers. Finally, Sentiment analysis was carried out on SO posts and comments for the top eleven cryptocurrencies to identify the prevailing sentiment. The sentiment analysis results showed that the sentiment of the SO posts and comments for all topics is predominantly *positive*. We observe that all the top cryptocurrencies have a similar distribution of sentiment across SO.

6.2 Future Work

In this study, we aim at understanding the use of software artifacts in cryptocurrency development. We obtain some interesting findings and have explained them. We now discuss several future directions that can be followed to improve our work.

First, the timeline of the GitHub’s panel dataset is nearly a year, it may require a longer term period to study development characteristics of cryptocurrencies to be able to extract more valuable insights from it. A new longitudinal dataset may reveal more accurate cryptocurrency price trend and volatility.

Many of the posts in SO are not properly tagged. Future research may consider a way for determining proper tags for SO discussions (e.g., auto-tagging or manual tagging). In our study we focused on extraction of key words from SO discussions. While in future, we could apply topic modelling to SO posts to be able to extract more relevant topics (e.g., cryptocurrency, Python, iOS, Javascript, etc.). We then can cluster various SO posts based on these topics and perhaps solving a problem of non-tagged posts.

Finally, we may extend our study by analyzing the popularity of cryptocurrency in SO for a specific time frame. For example, identifying the popular discussed cryptocurrencies in SO posts in the last 5 years and compare it with the results of popular discussed cryptocurrencies from the year 2010-2014. Such study can help us to understand how cryptocurrency popularity evolve over time. We can do the same for price volatility as well. In summary, there are several future directions that can be considered in extending the current work and improving the existing body of research.

List of References

- [1] R. van Tonder, A. Trockman, and C. Le Goues. “A Panel Data Set of Cryptocurrency Development Activity on GitHub.” In “International Conference on Mining Software Repositories,” MSR ’19 (2019).
- [2] A. Trockman, R. van Tonder, and B. Vasilescu. “Striking Gold in Software Repositories? An Econometric Study of Cryptocurrencies on GitHub.” In “International Conference on Mining Software Repositories,” MSR ’19 (2019).
- [3] N. Novielli, D. Girardi, and F. Lanubile. “A Benchmark Study on Sentiment Analysis for Software Engineering Research.” In “Proceedings of the 15th International Conference on Mining Software Repositories - MSR 18,” (2018).
- [4] E. Guzman, D. Azcar, and Y. Li. “Sentiment analysis of commit comments in github: an empirical study.” In “Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014, Hyderabad, India, May 31-June 01, 2014,” pages 352–355 (2014).
- [5] S. Baltes and S. Diehl. “Usage and attribution of Stack Overflow code snippets in GitHub projects.” *Empirical Software Engineering* **24**(3), 1259–1295 (2018).
- [6] S. Baltes, L. Dumani, C. Treude, and S. Diehl. “SOTorrent: reconstructing and analyzing the evolution of stack overflow posts.” In “Proceedings of the 15th International Conference on Mining Software Repositories, MSR 2018, Gothenburg, Sweden, May 28-29, 2018,” pages 319–330 (2018).
- [7] E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian. “The promises and perils of mining GitHub.” In “Proceedings of the 11th Working Conference on Mining Software Repositories, Hyderabad, India, May 31-June 01, 2014,” pages 92–101 (2014).
- [8] E. Kalliamvakou, D. Damian, K. Blincoe, L. Singer, and D. M. German. “Open source-style collaborative development practices in commercial projects using

- GitHub.” In “Proceedings of the 37th International Conference on Software Engineering - Volume 1, Florence, Italy, May 16-24, 2015,” pages 574–585 (2015).
- [9] G. Gousios and D. Spinellis. “GHTorrent: Githubs Data from a Firehose.” In “2012 9th IEEE Working Conference on Mining Software Repositories (MSR),” MSR ’12 (2012).
 - [10] G. Gousios. “The GHTorrent Dataset and Tool Suite.” In “2013 10th Working Conference on Mining Software Repositories (MSR),” MSR ’13 (2013).
 - [11] J. Tsay, L. Dabbish, and J. Herbsleb. “Let’s talk about it: evaluating contributions through discussion in GitHub.” In “Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering, Hong Kong, China, November 16-21, 2014,” pages 144–154 (2014).
 - [12] C. Treude and M. Wagner. “Predicting good configurations for GitHub and stack overflow topic models.” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 84–95 (2019).
 - [13] J. E. Montandon, L. L. Silva, and M. T. Valente. “Identifying experts in software libraries and frameworks among GitHub users.” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 276–287 (2019).
 - [14] S. S. Manes and O. Baysal. “How often and what StackOverflow posts do developers reference in their GitHub projects?” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 235–239 (2019).
 - [15] S. Wang, D. Lo, and L. Jiang. “An empirical study on developer interactions in stackoverflow.” In “Proceedings of the 28th Annual ACM Symposium on Applied Computing, Coimbra, Portugal, March 18-22, 2013,” pages 1019–1024 (2013).
 - [16] T. Diamantopoulos, M.-I. Sifaki, and A. L. Symeonidis. “Towards Mining Answer Edits to Extract Evolution Patterns in Stack Overflow.” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 215–219 (2019).
 - [17] A. Soni and S. Nadi. “Analyzing Comment-induced Updates on Stack Overflow.” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 220–224 (2019).

- [18] P. Chatterjee, K. Damevski, L. Pollock, V. Augustine, and N. A. Kraft. “Exploratory study of slack Q&A chats as a mining source for software engineering tools.” In “Proceedings of the 16th International Conference on Mining Software Repositories - MSR’19, Montreal, Quebec, Canada, May 26-27, 2019,” pages 490–501 (2019).
- [19] A. Bosu, C. S. Corley, D. Heaton, D. Chatterji, J. C. Carver, and N. A. Kraft. “Building reputation in StackOverflow: An empirical investigation.” In “Proceedings of the 10th Working Conference on Mining Software Repositories (MSR), San Francisco, CA, USA, May 18-19, 2013,” pages 89–92 (2013).
- [20] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra. “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis.” *SMU Data Science Review* **1**(3). Available at: <https://scholar.smu.edu/datasciencereview/vol1/iss3/1>, online; accessed on June 3, 2019 (2018).
- [21] M. Balcilar, E. Bouri, R. Gupta, and D. Roubaud. “Can volume predict Bitcoin returns and volatility? A quantiles-based approach.” *Economic Modelling* **64**, 74–81 (2017).
- [22] D. G. Baur, K. Hong, and A. D. Lee. “Bitcoin: Medium of exchange or speculative assets?” *Journal of International Financial Markets, Institutions and Money* **54**, 177–189 (2018).
- [23] L. Catania, S. Grassi, and F. Ravazzolo. “Forecasting Cryptocurrencies Financial Time Series.” *Centre for Applied Macro- and Petroleum economics (CAMP)* **1**(5). Available at: <https://biopen.bi.no/bi-xmlui/handle/11250/2489408>, online; accessed on June 10, 2019 (2018).
- [24] Y. Zhauniarovich, Y. Boshmaf, H. A. Jawaheri, and M. A. Sabah. “Characterizing Bitcoin donations to open source software on GitHub.” *arXiv preprint arXiv:1907.04002* (2019) <https://arxiv.org/pdf/1907.04002.pdf>, online; accessed on October 3, 2019 (2019).
- [25] A. Mittal, V. Dhiman, A. Singh, and C. Prakash. “Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data.” In “Proceedings of 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, August 08-09, 2019,” (2019).
- [26] A. Burnie and E. Yilmaz. “An Analysis of the Change in Discussions on Social Media with Bitcoin Price.” In “Proceedings of the 42nd International ACM

- SIGIR Conference on Research and Development in Information Retrieval, Paris, France, July 21-25, 2019,” pages 889–892 (2019).
- [27] M. R. Wrobel. “Emotions in the software development process.” In “Proceedings of the 6th International Conference on Human System Interactions (HSI), Sopot, Poland, June 06-08, 2013,” pages 518–523 (2013).
 - [28] F. Calefato, F. Lanubile, and N. Novielli. “How to ask for technical help? Evidence-based guidelines for writing questions on Stack Overflow.” *Information and Software Technology* **94**, 186–207 (2018).
 - [29] F. Jurado and P. Rodriguez. “Sentiment Analysis in monitoring software development processes: An exploratory case study on GitHubs project issues.” *Journal of Systems and Software* **104**(C), 82–89 (2015).
 - [30] D. Pletea, B. Vasilescu, and A. Serebrenik. “Security and emotion: sentiment analysis of security discussions on GitHub.” In “Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014, Hyderabad, India, May 31-June 01, 2014,” pages 348–351 (2014).
 - [31] M. Ortu, A. Murgia, G. Destefanis, P. Tourani, R. Tonelli, M. Marchesi, and B. Adams. “The emotional side of software developers in JIRA.” In “Proceedings of the 13th International Workshop on Mining Software Repositories - MSR 16, Austin, Texas, May 14-22, 2016,” pages 480–483 (2016).
 - [32] M. Ortu, G. Destefanis, B. Adams, A. Murgia, M. Marchesi, and R. Tonelli. “The JIRA Repository Dataset: Understanding Social Aspects of Software Development.” In “Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering, Beijing, China, October 21-21, 2015,” (2015).
 - [33] N. Novielli, D. Girardi, and F. Lanubile. “A benchmark study on sentiment analysis for software engineering research.” In “Proceedings of the 15th International Conference on Mining Software Repositories - MSR’18, Gothenburg, Sweden, May 28-29, 2018,” pages 364–375 (2018).
 - [34] E. Haddi, X. Liu, and Y. Shi. “The Role of Text Pre-processing in Sentiment Analysis.” *Procedia Computer Science* **17**, 26–32 (2013).
 - [35] T. Singh and M. Kumari. “Role of Text Pre-processing in Twitter Sentiment Analysis.” *Procedia Computer Science* **89**, 549–554 (2016).

- [36] C. Eom, T. Kaizoji, S. H. Kang, and L. Pichl. “Bitcoin and investor sentiment: Statistical characteristics and predictability.” *Physica A: Statistical Mechanics and its Applications* **514**, 511–521 (2019).
- [37] Y. Liu, Z. Qin, P. Li, and T. Wan. “Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis.” *Advances in Artificial Intelligence: From Theory to Practice Lecture Notes in Computer Science* **10350**, 192–201 (2017).
- [38] S. Baltes. *The SOTorrent Dataset*. <https://empirical-software-engineering/projects/sotorrent/>, online; accessed on July 2, 2019 (2019).
- [39] S. Baltes, C. Treude, and S. Diehl. “SOTorrent: Studying the Origin, Evolution, and Usage of Stack Overflow Code Snippets.” In “Proceedings of the 16th International Conference on Mining Software Repositories, MSR 2019, Montreal, Quebec, May 26-27, 2019,” (2019).
- [40] P. J. Brockwell and R. A. Davis. “Introduction to Time Series and Forecasting, second edition.” *Springer* (2002).
- [41] D. A. Dickey and W. A. Fuller. “Distribution of the Estimators for Autoregressive Time Series With a Unit Root.” *Journal of the American Statistical Association* **74**, 427–431 (1979).
- [42] *AnalyticsVidhya*. <https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python/>, online; accessed on October 3, 2019 (2019).
- [43] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” *Journal of Econometrics* **54**, 159–178 (1992).
- [44] H. Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle.” *Springer Series in Statistics, Selected Papers of Hirotugu Akaike* pages 199–213 (1998).
- [45] C. W. Granger. “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods.” *Econometrica: Journal of the Econometric Society* **37**, 424–438 (1969).
- [46] J. M. Wooldridge. “Introductory Econometrics: A Modern Approach.” *Nelson Education* (2015).

- [47] B. Liu. “Sentiment Analysis and Opinion Mining.” *Synthesis lectures on human language technologies* **5**, 1–167 (2012).

Appendix A

Ranking Popular Repositories of Popular Cryptocurrencies in GitHub

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
go-ethereum	1,327
wiki	985
mist	614
EIPs	486
cpp-ethereum	461
solidity	340
pyethereum	273
web3.js	242
ethereumj	202
research	151

Table 43: Popular repositories of **Ethereum** based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
go-ethereum	2,041
wiki	1,300
EIPs	829
mist	769
aleth	511
solidity	496
web3.js	399
pyethereum	284
ethereumj	247
research	220

Table 44: Popular repositories of **Ethereum** based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
go-ethereum	12,338
wiki	7,131
mist	5,087
web3.js	3,483
solidity	3,186
cpp-ethereum	2,370
pyethereum	2,013
EIPs	1,812
vyper	1,199
ethereumj	1,090

Table 45: Popular repositories of **Ethereum** based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
go-ethereum	22,492
wiki	11,961
web3.js	7,499
mist	7,221
solidity	6,610
EIPs	4,258
aleth	3,491
pyethereum	2,646
vyper	2,217
ethereumj	1,907

Table 46: Popular repositories of **Ethereum** based on stars on 2019-02-04.

Ethereum	
No.	Repository
1	go-ethereum
2	wiki
3	web3.js
4	mist
5	solidity
6	EIPs
7	pyethereum
8	ethereumj

Table 47: Popular repositories of **Ethereum** in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
ripple-client	730
rippled	420
ripple-lib	158
ilp-plugin-xrp-escrow	62
ilp-plugin-xrp-paychan	59
ripplecharts-frontend	58
ripple-client-desktop	52
ripple-dev-portal	37
ripple-lib-java	34
rippled-historical-	
-database	33

Table 48: Popular repositories of **Ripple** based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
ripple-client	724
rippled	473
ripple-lib	184
ripplecharts-frontend	64
ripple-client-desktop	55
ripple-dev-portal	47
rippled-historical-	
-database	37
ripple-data-api	31
ripple-blobvault	30
ripple-rest	27

Table 49: Popular repositories of **Ripple** based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
rippled	2,371
ripple-client	1,327
ripple-lib	604
ripplecharts-frontend	138
ripple-client-desktop	100
ripple-dev-portal	84
ripple-lib-java	81
ripple-rest	73
ripple-data-api	45
rippled-historical-	
-database	42

Table 50: Popular repositories of **Ripple** based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
rippled	3,127
ripple-client	1,357
ripple-lib	790
ripplecharts-frontend	159
ripple-dev-portal	133
ripple-client-desktop	114
ripple-rest	79
rippled-historical-	
-database	70
ripple-data-api	45
ripple-dot-net	42

Table 51: Popular repositories of **Ripple** based on stars on 2019-02-04.

Ripple	
No.	Repository
1	rippled
2	ripple-client
3	ripple-lib
4	ripplecharts-frontend
5	ripple-dev-portal
6	ripple-client-desktop
7	rippled-historical-database
8	ripple-rest
9	ripple-data-api

Table 52: Popular repositories of Ripple in GitHub.

Popularity (Watchers) on 2018-01-21		Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers	Repository	# of Watchers
cardano-sl	278	cardano-sl	408
Scorex	68	daedalus	93
daedalus	66	dev-handbook	67
rscoin-haskell	39	Scorex	65
plutus-prototype	34	plutus	62
stack2nix	34	cardano-mainnet-mirror	60
rscoin-core	25	stack2nix	56
cardano-sl-style-		cardano-chain	56
-guides	25	cardano-prelude	52
rscoin-nixops	25	iohk-monitoring-	
cardanodocs.com	24	-framework	51

Table 53: Popular repositories of Cardano based on watchers on 2018-01-21.

Table 54: Popular repositories of Cardano based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
cardano-sl	1,690
Scorex	448
daedalus	268
scrypto	98
iodb	72
mantis	64
plutus-prototype	51
iohk-ops	46
stack2nix	40
rscoin-haskell	37

Table 55: Popular repositories of **Cardano** based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
cardano-sl	3,146
daedalus	526
Scorex	484
plutus	178
mantis	158
scrypto	154
rust-cardano	137
iodb	88
iohk-ops	86
stack2nix	71

Table 56: Popular repositories of **Cardano** based on stars on 2019-02-04.

Cardano	
No.	Repository
1	cardano-sl
2	daedalus
3	Scorex
4	stack2nix

Table 57: Popular repositories of **Cardano** in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
litecoin	522
lips	55
liteaddress.org-old	33
litecoin.org-ancient	32
litecore-litecoin	30
litecoin.org	29
p2pool-scanner	29
litecore	26
litecore-lib	22
Electrum-obsolete	20

Table 58: Popular repositories of Litecoin based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
litecoin	539
lips	52
litecore-litecoin	43
liteaddress.org-old	32
litecoin.org-ancient	31
old-litecoin-org	28
p2pool-scanner	27
litecore	20
litecore-lib	19
Electrum-obsolete	19

Table 59: Popular repositories of Litecoin based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
litecoin	2,674
liteaddress.org	61
liteaddress.org-old	42
litecoin.org-ancient	39
lips	37
litecore-litecoin	36
litecore	32
litecoin.org	29
litecore-lib	24
p2pool-scanner	24

Table 60: Popular repositories of Litecoin based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
litecoin	3,242
liteaddress.org	76
litecore-litecoin	76
litecore-lib	46
liteaddress.org-old	41
lips	40
litecore	40
litecoin.org-ancient	39
litecore-node	29
old-litecoin-org	26

Table 61: Popular repositories of Litecoin based on stars on 2019-02-04.

Litecoin	
No.	Repository
1	litecoin
2	litecore-litecoin
3	litecore-lib
4	liteaddress.org-old
5	lips
6	litecore
7	litecoin.org-ancient
8	p2pool-scanner

Table 62: Popular repositories of Litecoin in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
eos	722
Documentation	195
eos-token-distribution	43
chainbase	26
genesis	26
eosjs	25
eos-wallet-app	25
appbase	20
logchain	17
eosjs-api	16

Table 63: Popular repositories of EOS based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
eos	1,293
Documentation	322
eosjs	121
eosio.contracts	64
eosio.cdt	61
demux-js	59
eosjs-api	54
eosjs-json	48
eosjs-ecc	46
eosjs-keygen	46

Table 64: Popular repositories of EOS based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
eos	3,569
Documentation	819
eos-token-	
-distribution	137
eosjs	106
eos-wallet-app	48
chainbase	38
appbase	19
eosjs-api	15
genesis	12
eosjs-ecc	10

Table 65: Popular repositories of EOS based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
eos	10,050
Documentation	2,013
eosjs	1,097
eosio.cdt	307
demux-js	232
eos-token-distribution	203
eosjs-ecc	183
eosjs-api	158
eosio.contracts	154
eosio-project-	
-boilerplate-simple	113

Table 66: Popular repositories of EOS based on stars on 2019-02-04.

EOS	
No.	Repository
1	eos
2	Documentation
3	eosjs
4	eosjs-api
5	eos-token-distribution
6	eosjs-ecc

Table 67: Popular repositories of EOS in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
stellar-core	183
stellard	68
docs	54
horizon	50
stellar-client	48
js-stellar-sdk	44
stellar-protocol	42
go	40
stellar-wallet	40
stellar-viewer	33

Table 68: Popular repositories of Stellar based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
stellar-core	261
go	68
stellard	67
stellar-protocol	67
horizon	65
js-stellar-sdk	64
docs	60
stellar-client	50
stellar-wallet	46
bridge-server	38

Table 69: Popular repositories of Stellar based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
stellar-core	1,060
stellar-client	308
stellard	270
js-stellar-sdk	153
horizon	125
go	123
docs	93
stellar-wallet	90
dashboard	55
stellar-lib	55

Table 70: Popular repositories of Stellar based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
stellar-core	1,983
go	439
js-stellar-sdk	341
stellar-client	312
stellard	278
horizon	213
stellar-protocol	142
docs	141
dashboard	108
stellar-wallet	99

Table 71: Popular repositories of Stellar based on stars on 2019-02-04.

Stellar	
No.	Repository
1	stellar-core
2	go
3	js-stellar-sdk
4	stellar-client
5	stellard
6	horizon
7	docs
8	stellar-wallet
9	stellar-protocol

Table 72: Popular repositories of Stellar in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
wallet	276
iri	167
iota.lib.js	82
iota.lib.py	57
iota.lib.java	23
iota.lib.csharp	22
iota.lib.go	21
iota.rs	20
android-wallet-app	18
MAM	17

Table 73: Popular repositories of IOTA based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
wallet	289
iri	168
iota.js	108
iota.lib.py	66
trinity-wallet	36
entangled	30
iota.lib.csharp	30
iota.go	27
iota-java	27
MAM	22

Table 74: Popular repositories of IOTA based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
wallet	1,872
iri	735
iota.lib.js	470
iota.lib.py	192
iota.lib.go	93
android-wallet-app	73
iota.rs	63
iota.lib.java	62
iota.flash.js	56
wiki	47

Table 75: Popular repositories of IOTA based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
wallet	2,168
iri	1,111
iota.js	727
iota.lib.py	292
trinity-wallet	224
iota.go	174
iota-java	105
android-wallet-app	84
iota.rs	77
iota.flash.js	66

Table 76: Popular repositories of IOTA based on stars on 2019-02-04.

IOTA	
No.	Repository
1	wallet
2	iri
3	iota.lib.py
4	mist
5	android-wallet-app
6	iota.rs

Table 77: Popular repositories of IOTA in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
monero	350
kovri	88
monero-gui	68
research-lab	39
monero-site	38
meta	32
sekura	12
kovri-site	11
monero-forum	11
mininero	9

Table 78: Popular repositories of Monero based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
monero	456
kovri	98
monero-gui	86
research-lab	47
meta	43
monero-site	40
kastelo	18
kovri-docs	13
kovri-site	12
monero-forum	11

Table 79: Popular repositories of Monero based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
monero	2,312
kovri	418
monero-gui	243
research-lab	56
monero-site	47
meta	33
urs	30
sekura	20
mininero	19
kovri-docs	18

Table 80: Popular repositories of Monero based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
monero	3,602
kovri	548
monero-gui	387
research-lab	89
monero-site	77
meta	55
urs	49
kastelo	36
mininero	36
kovri-docs	22

Table 81: Popular repositories of Monero based on stars on 2019-02-04.

Monero	
No.	Repository
1	monero
2	kovri
3	monero-gui
4	research-lab
5	monero-site
6	meta
7	mininero
8	kovri-docs

Table 82: Popular repositories of Monero in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
neo	355
neo-gui	47
docs	33
proposals	27
examples-csharp	26
neo-vm	19
neo.org	17
neo-cli	16
leveldb	13
examples-java	12

Table 83: Popular repositories of NEO based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
neo	422
neo-gui	78
proposals	42
docs	38
neo-cli	38
examples-csharp	36
neo-vm	36
examples-java	21
neo-compiler	20
neo-devpack-dotnet	18

Table 84: Popular repositories of NEO based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
neo	1,679
neo-gui	160
docs	121
neo-cli	69
examples-csharp	65
proposals	56
neo-vm	38
neo.org	30
neo-compiler	28
neo-devpack-dotnet	24

Table 85: Popular repositories of NEO based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
neo	2,742
neo-gui	223
docs	169
neo-cli	141
examples-csharp	120
proposals	106
neo-vm	97
neo-compiler	55
neo.org	40
neo-devpack-dotnet	33

Table 86: Popular repositories of NEO based on stars on 2019-02-04.

NEO	
No.	Repository
1	neo
2	neo-gui
3	docs
4	neo-cli
5	examples-csharp
6	proposals
7	neo-vm
8	neo-compiler
9	neo.org
10	neo-devpack-dotnet

Table 87: Popular repositories of NEO in GitHub.

Popularity (Watchers) on 2018-01-21	
Repository	# of Watchers
status-react	187
status-go	77
open-bounty	49
wiki.status.im	47
ideas	42
contracts	42
react-native-desktop	41
status-lib	41
status-network-token	40
docs.status.im	40

Table 88: Popular repositories of **Status** based on watchers on 2018-01-21.

Popularity (Watchers) on 2019-02-04	
Repository	# of Watchers
status-react	189
status-go	79
nimbus	56
contracts	53
wiki.status.im	52
react-native-desktop	50
swarms	50
open-bounty	47
status-lib	45
web3.js	44

Table 89: Popular repositories of **Status** based on watchers on 2019-02-04.

Popularity (Stars) on 2018-01-21	
Repository	# of Stars
status-react	1,821
status-go	301
status-network-token	97
open-bounty	94
react-native-desktop	89
truffle-box-status	34
hackathon	29
hardware-wallet	26
wiki.status.im	25
ideas	23

Table 90: Popular repositories of **Status** based on stars on 2018-01-21.

Popularity (Stars) on 2019-02-04	
Repository	# of Stars
status-react	2,549
react-native-desktop	563
status-go	479
nimbus	211
status-network-token	126
open-bounty	123
swarms	78
contracts	72
status-keycard	58
clj-rn	48

Table 91: Popular repositories of **Status** based on stars on 2019-02-04.

Status	
No.	Repository
1	status-react
2	react-native-desktop
3	status-go
4	open-bounty
5	status-network-token
6	contracts
7	wiki.status.im

Table 92: Popular repositories of **Status** in GitHub.

Appendix B

Development Activities of Popular Repositories of Popular Cryptocurrencies in GitHub

(a) Watchers

(b) Stars

Figure 24: Trends in popularity in popular repositories of Ethereum.

(a) Forks

(b) Weekly Contributors

Figure 25: Trends in activity in popular repositories of Ethereum.

(a) Daily

(b) Weekly

(c) Yearly

Figure 26: Trends in commits in popular repositories of Ethereum.

(a) Daily

(b) Weekly

(c) Yearly

Figure 27: Trends in LOCs added in popular repositories of Ethereum.

(a) Daily

(b) Weekly

(c) Yearly

Figure 28: Trends in LOCs removed in popular repositories of Ethereum.

(a) Watchers

(b) Stars

Figure 29: Trends in popularity in popular repositories of Ripple.

(a) Forks

(b) Weekly Contributors

Figure 30: Trends in activity in popular repositories of Ripple.

(a) Daily

(b) Weekly

(c) Yearly

Figure 31: Trends in commits in popular repositories of Ripple.

(a) Daily

(b) Weekly

(c) Yearly

Figure 32: Trends in LOCs added in popular repositories of Ripple.

(a) Daily

(b) Weekly

(c) Yearly

Figure 33: Trends in LOCs removed in popular repositories of Ripple.

(a) Watchers

(b) Stars

Figure 34: Trends in popularity in popular repositories of Cardano.

(a) Forks

(b) Weekly Contributors

Figure 35: Trends in activity in popular repositories of Cardano.

(a) Daily

(b) Weekly

(c) Yearly

Figure 36: Trends in commits in popular repositories of Cardano.

(a) Daily

(b) Weekly

(c) Yearly

Figure 37: Trends in LOCs added in popular repositories of Cardano.

(a) Daily

(b) Weekly

(c) Yearly

Figure 38: Trends in LOCs removed in popular repositories of Cardano.

(a) Watchers

(b) Stars

Figure 39: Trends in popularity in popular repositories of Litecoin.

(a) Forks

(b) Weekly Contributors

Figure 40: Trends in activity in popular repositories of Litecoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 41: Trends in commits in popular repositories of Litecoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 42: Trends in LOCs added in popular repositories of Litecoin.

(a) Daily

(b) Weekly

(c) Yearly

Figure 43: Trends in LOCs removed in popular repositories of Litecoin.

(a) Watchers

(b) Stars

Figure 44: Trends in popularity in popular repositories of EOS.

(a) Forks

(b) Weekly Contributors

Figure 45: Trends in activity in popular repositories of EOS.

(a) Daily

(b) Weekly

(c) Yearly

Figure 46: Trends in commits in popular repositories of EOS.

(a) Daily

(b) Weekly

(c) Yearly

Figure 47: Trends in LOCs added in popular repositories of EOS.

(a) Daily

(b) Weekly

(c) Yearly

Figure 48: Trends in LOCs removed in popular repositories of EOS.

(a) Watchers

(b) Stars

Figure 49: Trends in popularity in popular repositories of Stellar.

(a) Forks

(b) Weekly Contributors

Figure 50: Trends in activity in popular repositories of Stellar.

(a) Daily

(b) Weekly

(c) Yearly

Figure 51: Trends in commits in popular repositories of Stellar.

(a) Daily

(b) Weekly

(c) Yearly

Figure 52: Trends in LOCs added in popular repositories of Stellar.

(a) Daily

(b) Weekly

(c) Yearly

Figure 53: Trends in LOCs removed in popular repositories of Stellar.

(a) Watchers

(b) Stars

Figure 54: Trends in popularity in popular repositories of IOTA.

(a) Forks

(b) Weekly Contributors

Figure 55: Trends in activity in popular repositories of IOTA.

(a) Daily

(b) Weekly

(c) Yearly

Figure 56: Trends in commits in popular repositories of IOTA.

(a) Daily

(b) Weekly

(c) Yearly

Figure 57: Trends in LOCs added in popular repositories of IOTA.

(a) Daily

(b) Weekly

(c) Yearly

Figure 58: Trends in LOCs removed in popular repositories of IOTA.

(a) Watchers

(b) Stars

Figure 59: Trends in popularity in popular repositories of Monero.

(a) Forks

(b) Weekly Contributors

Figure 60: Trends in activity in popular repositories of Monero.

(a) Daily

(b) Weekly

(c) Yearly

Figure 61: Trends in commits in popular repositories of Monero.

(a) Daily

(b) Weekly

(c) Yearly

Figure 62: Trends in LOCs added in popular repositories of Monero.

(a) Daily

(b) Weekly

(c) Yearly

Figure 63: Trends in LOCs removed in popular repositories of Monero.

(a) Watchers

(b) Stars

Figure 64: Trends in popularity in popular repositories of NEO.

(a) Forks

(b) Weekly Contributors

Figure 65: Trends in activity in popular repositories of NEO.

(a) Daily

(b) Weekly

(c) Yearly

Figure 66: Trends in commits in popular repositories of NEO.

(a) Daily

(b) Weekly

(c) Yearly

Figure 67: Trends in LOCs added in popular repositories of NEO.

(a) Daily

(b) Weekly

(c) Yearly

Figure 68: Trends in LOCs removed in popular repositories of NEO.

(a) Watchers

(b) Stars

Figure 69: Trends in popularity in popular repositories of Status.

(a) Forks

(b) Weekly Contributors

Figure 70: Trends in activity in popular repositories of Status.

(a) Daily

(b) Weekly

(c) Yearly

Figure 71: Trends in commits in popular repositories of Status.

(a) Daily

(b) Weekly

(c) Yearly

Figure 72: Trends in LOCs added in popular repositories of Status.

(a) Daily

(b) Weekly

(c) Yearly

Figure 73: Trends in LOCs removed in popular repositories of Status.

Word Clouds of Hot topics in SO Discussions Related to Cryptocurrency

Appendix D

Sentiment Distributions in SO Comments Related to Cryptocurrency

Figure 81: Distribution of sentiment in comments of SO posts related to Bitcoin.

Figure 82: Distribution of sentiment in comments of SO posts related to Ethereum.

Figure 83: Distribution of sentiment in comments of SO posts related to Ripple.

Figure 84: Distribution of sentiment in comments of SO posts related to EOS.

Figure 85: Distribution of sentiment in comments of SO posts related to Status.

Figure 86: Distribution of sentiment in comments of SO posts related to Blockchain tag.

Figure 87: Distribution of sentiment in comments of SO posts related to Cryptocurrency tag.

Topic	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
Blockchain	207	942	144	615	464	692	271	1,334	1,065	2,362
Cryptocurrency	14	49	3	31	26	46	21	81	78	173
Bitcoin	186	617	110	432	352	519	221	884	865	1,770
Ethereum	175	800	148	508	432	640	251	1,320	970	1,948
Ripple	34	138	36	154	108	171	59	218	249	405
EOS	3	13	2	6	5	10	1	17	17	21
Status	149	503	118	402	331	454	184	655	670	1,588
All Posts	2,500	11,825	1,698	6,462	6,220	8,125	3,112	14,604	12,147	26,883
All Comments	704	2,755	518	1,958	1,566	2,329	940	4,058	3,576	7,513

Table 93: Sentiment Scores.