

Characterizing Bitcoin donations to open source software on GitHub

Yury Zhauniarovich*
Perfect Equanimity

Husam Al Jawaheri*
University of Luxembourg

Yazan Boshmaf
Qatar Computing Research Institute, HBKU

Mashaël Al Sabah
Qatar Computing Research Institute, HBKU

ABSTRACT

Web-based hosting services for version control, such as GitHub, have made it easier for people to develop, share, and donate money to software repositories. In this paper, we study the use of Bitcoin to make donations to open source repositories on GitHub. In particular, we analyze the amount and volume of donations over time, in addition to its relationship to the age and popularity of a repository.

We scanned over three million repositories looking for donation addresses. We then extracted and analyzed their transactions from Bitcoin’s public blockchain. Overall, we found 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. We also found 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.

1 INTRODUCTION

Open source software has revolutionized our world. Developers around the world can now explore open source repositories, learn best practices, and collaborate on improving software used by millions of users. Web-based hosting services for version control, such as GitHub [3], GitLab [4], and Bitbucket [1], have played an important role in this revolution. Developers today use these services to facilitate all stages of software engineering, including planning, version control, issue tracking, release management, and software deployment. Among these platforms, GitHub stands out as the most popular with more than 96 million repositories [20], out of which more than three million are open sourced under MIT, Apache, GPL, or other open source software licenses [19].

As open source repositories require considerable time and money investments by developers, it is customary to ask users for financial aid, often in the form of donations, to support and maintain their development. While there are several platforms that facilitate this process [26], they typically charge a relatively high transaction fee that makes them infeasible for developers, given the small amount and volume of donations. Moreover, such platforms are not available to everyone because of country-specific legal restrictions or limited access to financial institutions. To overcome these challenges, developers started accepting cryptocurrencies as a payment method for receiving donations. Among many cryptocurrencies, Bitcoin is the most popular with a multi-billion dollar market.¹

While Bitcoin and GitHub have co-existed for 10 years, we still do not know how prominent Bitcoin donations are, specifically to open source repositories, and how repository-specific factors, such as popularity metrics, affect the amount and volume of donations.

In this paper, we bridge this knowledge gap and analyze the use of Bitcoin to make donations to open source repositories on GitHub.

We used Google BigQuery to scan over three million open source repositories looking for Bitcoin addresses, which are explicitly listed in “readme” files by repository owners for receiving donations (§3). We then manually reviewed and validated the collected addresses, ending up with 1,996 donation addresses associated with 6,075 repositories. We also developed a web crawler and used it to directly fetch repository-specific information from GitHub, such as repository creation date, whether it is a fork, and popularity metrics like number of watchers, stars, and forks.

To analyze Bitcoin donations, we first extracted all transactions that involve any of the collected donation addresses from Bitcoin’s public blockchain, ending up with 56,454 transactions. We then performed transaction and correlation analyses (§4). Our results show that there is a limited adoption of Bitcoin as a payment method for receiving donations, and that the amount of donations does not strongly correlate with the popularity of a repository. We found that only 0.2% of the repositories use Bitcoin for receiving donation, and that for those which do, only 51.8% of their donation addresses have ever received deposits. In fact, just 9.2 thousand bitcoins have been donated to open source repositories on GitHub over the last 10 years, which is equivalent to 8.3 million dollars, calculated retrospectively. Moreover, the top-10 largest donations in value, which constituted 70.23% of all donations in dollars, were made between late 2017 and early 2018 when the price of Bitcoin reached its record high. As such, this speculative and deflationary nature of bitcoin price could explain the disproportionately small amount of donations, where people are discouraged from making donations hoping for a higher future price. Another reason behind limited adoption of Bitcoin by repository owners is privacy, as once an address is publicly listed it could be linked to its owner and traced to personal transactions [12].

We also found weak positive correlation between donation value in dollars and repository popularity, where the highest correlation ($r=0.013$) was associated with number of forks. Indeed, the top-10 grossing donation addresses, which received 54% of all donations, are associated with unpopular repositories. In particular, six of these addresses are associated with single repositories that each had at most two watchers, stars, or forks. To some extent, this counter intuitive result can be explained by the fact that some unpopular repositories list personal addresses, which might receive deposits for reasons other than donations, or donation addresses of well-known open source organizations, which attract larger amounts of donations from channels other than GitHub.

To this end, this paper contributes with the first donation study which depends on actual data collected from public software repositories and cryptocurrency blockchains, instead of collecting self-reported data through user surveys. Even though the results draw a

*Research was conducted while at Qatar Computing Research Institute, HBKU.

¹At the time of writing, the market cap for Bitcoin is more than 107 billion dollars [2].

Table 1: Bitcoin address types and regex patterns.

Type	Description	Regex pattern
P2PKH	Pay to public key hash	1[a-km-zA-HJ-NP-Z1-9]{25,34}
P2SH	Pay to script hash	3[a-km-zA-HJ-NP-Z1-9]{25,34}
Bech32	Segregated witness	bc1[a-zA-HJ-NP-Z0-9]{25,39}

grim image, we remind the reader that people and institutions just started to seriously consider Bitcoin as a currency or an asset class, and that recent considerable donations, such as the one million dollar transaction to the Free Software Foundation by the anonymous Pineapple Fund [16], are promising examples of how Bitcoin can be used to support the development of open source software.

2 BACKGROUND

We now present a brief background on Bitcoin and GitHub.

2.1 Bitcoin

Bitcoin is the first and the most popular cryptocurrency network [28]. In Bitcoin, the identity of a user is hidden by using public pseudonyms called addresses. A Bitcoin address is an alphanumeric identifier that is derived from the public key of a public/private key pair. It has three different formats in use today, each defining an address type, as specified by the regular expression patterns shown in Table 1.

The set of public/private keys that are owned by a user is called a wallet. Private keys are used to sign inputs of transactions as a proof of ownership. For example, if Alice wants to donate bitcoins to Bob, she creates a new transaction specifying one or more addresses from her wallet as inputs. She also specifies the amount to be transferred, as permitted by the amount previously received on the used inputs, and chooses one or more addresses from Bob’s wallet as transaction outputs. To protect the transaction, she signs it using her private keys, and then broadcasts it to the whole network for verification.

All Bitcoin transactions are stored in a decentralized ledger called a blockchain, which also means anyone could try to identify user transactions by analyzing the blockchain. This task, however, is typically difficult since user identities are not recorded in the blockchain, only Bitcoin addresses. As such, Bitcoin does not provide complete anonymity but rather pseudonymity. If user identities are linked to Bitcoin addresses, their transactions can be easily identified. This is the case when users publicly post their Bitcoin addresses online along with their personally identifiable information, such as posting personal donation addresses on public GitHub repositories.

It is possible to exchange bitcoins with fiat currencies using cryptocurrency exchanges. While largely unregulated, the price is often based on supply and demand, but can still vary significantly during a day. Services like CoinMarketCap [2] and Yahoo Finance [11] aggregate pricing data from exchanges and provide market statistics, such as open/close and high/low values for a trading day.

2.2 GitHub

With more than 31 million developers worldwide [20], GitHub is one of the most popular web-based hosting services for Git version control. The service allows developers to upload their software

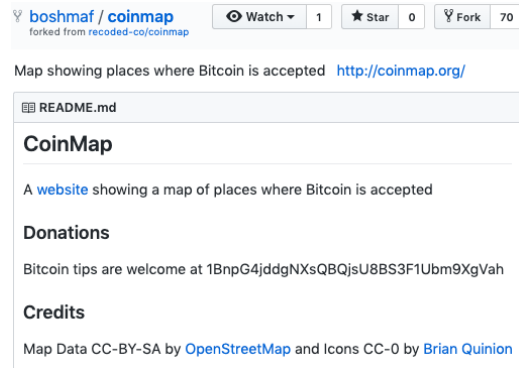


Figure 1: Repository information on GitHub.

projects into code repositories, which enables them to display, review, search, and navigate through their source code. Moreover, the service provides issue tracking, project planning, documentation, and release management tools. There is also a social networking feature that allows developers to watch, star, and fork repositories. As such, these social interactions can be used to estimate the popularity of a repository, as more popular repositories are likely to have a larger number of watchers, stars, and forks than others.

On GitHub, a user can create a new repository, either public or private, or fork (i.e., copy) an existing public repository. There are more than 96 million repositories hosted on GitHub, out of which 30 million are public and 3.4 million are open sourced under MIT, Apache, GPL, or other open source licenses [19, 20]. Each repository has an identifier that consists of two parts: the repository’s owner, which is the developer’s alias on GitHub, and the repository’s name. When forking a repository, the repository’s name remains the same but the owner is changed to the forking developer. Figure 1 shows an example of how repository information is displayed on GitHub. From this figure, we know a repository called “coinmap” has been forked from a repository with the identifier “recoded-co/coinmap” by a developer with the alias “boshmaf.”

We note that the displayed number of forks, also called network size, always refers to the root repository. For example, the number shown in Figure 1 means “recoded-co” had their repository forked 70 times, not “boshmaf.” However, using GitHub’s API, it is possible to find out the number of immediate forks of a forked repository. To distinguish between these two numbers, we refer to the earlier as the network size and the latter as the number of forks.

In addition to repository information, GitHub also parses and displays text files called “readme” that are often used by developers to describe their work. By default, the service scans the “root,” “docs” and “.github” directories of a repository for such files [18].

3 DATASET

We now describe the dataset we used in this study.

3.1 Summary

Overall, we collected 1,996 Bitcoin donation addresses associated with 6,075 open source repositories on GitHub. In terms of address types, P2PKH was the most popular with 1,843 addresses (92.33%),

followed by P2SH with 144 addresses (7.21%), and finally the newly introduced Bech32 with 9 addresses (0.45%).

As for the repositories, a total of 2,095 repositories (34.49%) were forks. Overall, the repositories belonged to 4,686 owners, out of which 434 owners (9.26%) represented organizations.

3.2 Collection

Our goal is to collect public information about open source repositories on GitHub that accept Bitcoin donations. As there are more than 96 million repositories hosted on GitHub [20], it is infeasible to download and inspect all of their contents. Instead, we used a recent snapshot of GitHub’s open source repositories that is hosted on Google BigQuery service, which is available free of charge as part of Google Cloud Public Dataset Program [21, 22]. We extracted a dataset of repositories that contain one or more Bitcoin addresses in their “readme” files. The snapshot we used in this study was updated on Dec 5, 2018, and includes files, commits, licences, and language information of over three million repositories [19].

We decided to search “readme” files because developers usually include a donation request as part of the overall repository description. In particular, we search “root,” “docs,” and “.github” directories of each repository for files whose name contains “readme” substring (e.g., “README.md”). We then scan the content of each identified file for Bitcoin addresses using regular expressions, as specified in Table 1. Google provides a possibility to search over the content of non-binary files on the “HEAD” branch that are less than 1MB [22].

The snapshot we used does not include repository-specific information, such as its creation date, whether it is a fork or owned by an organization, and popularity-related information. To overcome this limitation, we developed a simple Python crawler to retrieve the missing information from Github, and download latest versions of “readme” files that contain Bitcoin addresses. We ran the crawler on Dec 26, 2018, and it took about 12 hours to update the dataset.

To this end, the collected dataset consisted of 7,968 repositories. Overall, there were 6,632 unique owners and 3,861 unique Bitcoin addresses, which means some owners have several repositories, and some addresses are used in multiple repositories.

3.3 Cleaning

Some of the collected addresses might be invalid Bitcoin addresses. For example, an address that consists of all 1s or 3s is invalid. Moreover, even if an address is valid, it could be the case it is listed in the “readme” file for reasons others than accepting donations, such as showing a coding example that uses a well-known Bitcoin address.

In order to clean the dataset, we manually reviewed and validated the addresses associated with each repository against their use in “readme” files. We only kept addresses that are valid Bitcoin addresses which are used in a textual donation request, a donation link or badge, or the signature of a repository’s owner. For example, Figure 1 shows an address that matches this criteria. Out of 3,861 addresses associated with 7,968 repositories, we ended up with 1,996 valid Bitcoin donation addresses associated with 6,075 repositories.

4 ANALYSIS

We now present the analysis we performed on the collected dataset.

4.1 Summary

Overall, we found 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. We also found weak positive correlation between the amount of donation in dollars and the popularity of a repository, with highest correlation ($r=0.013$) associated with number of forks.

4.2 Donations

Our goal is to analyze the amount and volume of Bitcoin donations. To achieve this, we deployed BlockSci [25] on Jan 2, 2019, and used it to extract all transactions from Bitcoin’s public blockchain which include one or more donation addresses as inputs or outputs, ending up with 56,454 transactions. If a transaction increases the balance of an address, we call it a deposit. Otherwise, we call the transaction a withdrawal. This classification is important when an address is used as both an input and an output of a transaction. We calculate the total deposit or withdrawal value for each donation address as the sum of increments or decrements of its balance in dollars, based on the exchange rate at the time of each corresponding transaction. As there is no evident way to exclude deposits to an address that are not donations, we refer to the total deposit value and the number of deposits as the amount and volume of donation, respectively. This means the results we report herein are optimistic and represent upper-bound estimates of the real amount and volume of donations.

Overall, a total of 8,348,241 dollars ($\approx 9,238.9516$ bitcoins) were received through 43,862 deposits, and 25,889,923 dollars ($\approx 9,015.6384$ bitcoins) were sent through 12,592 withdrawals. On average, a donation address received 4,182 dollars (≈ 4.6287 bitcoins) through 21.97 deposits and sent 4,196 dollars (≈ 4.5169 bitcoins) through 6.31 withdrawals. This money flow, and an average balance of 0.0958 bitcoins, show that deposits to donation addresses are rarely accumulated, but are rather spent or sent to other addresses as withdrawals.

In general, deposits and withdrawals are highly skewed across donation addresses. Only 1,034 addresses (51.8%) have ever received deposits, with the top-10 grossing addresses receiving 70.23% of all donations in dollars, or 37.44% in bitcoins, as shown in Table 2. Similarly, only 873 addresses (43.74%) have ever been used in withdrawals. With an average repository age of 3.49 years, these findings suggest that Bitcoin has a limited adoption as a payment method for sending or receiving donations.²

4.2.1 Top-grossing addresses. While the average donation amount and volume per address is relatively low, a few addresses were clear outliers, as shown in Table 2. In what follows, we manually inspect these addresses and the associated repositories looking for clues that could explain their success in receiving donations.

First, donation addresses associated with older and more popular repositories do not necessarily receive larger amounts of donations. While on average a donation address has 25.85 watchers, 281.87 stars, 75.85 forks, and a network size of 3,009.37 from nearly three repositories over 3.49 years, most of the addresses in Table 2 are associated with repositories that fall below these values. Also, all but one of the top-10 addresses which are listed in the most popular

²We could not measure the time at which a donation address is listed in a “readme” file, so we report the time at which the associated repository was created (i.e., its age).

Table 2: Top-10 grossing donation addresses in dollars.

Address	Tag?	Org?	# repos	Repositories						Donations		
				Avg. age (years)	Total popularity			Net. size	Amount		Volume (txes)	
					# watchers	# stars	# forks		(bitcoins)	(dollars)		
1GbAWBiGyu...	✗	✗	1	1.98	1	1	9	9	364.9295	1,689,358	113	
1257U991Wq...	✗	✗	1	2.55	1	0	0	0	1311.7105	1,302,104	144	
1PC9aZC4hN...	✓	✓	2	2.98	71	825	513	513	1085.1980	1,224,570	1772	
33ENWZ9RCY...	✓	✗	2	1.81	60	213	234	234	102.9535	365,765	25	
1MEWT2SGbq...	✗	✗	2	1.29	2	0	0	1602	80.7299	331,126	206	
1BsTwoMaX3...	✗	✗	1	3.29	1	0	0	151	284.1289	258,347	228	
1FFPahYGPH...	✗	✗	1	1.52	3	4	4	4	103.4162	220,196	95	
18xgGTTzZU...	✗	✗	1	3.80	1	0	0	0	46.9061	168,378	65	
3DKLj6rEZN...	✗	✓	1	2.59	22	158	11	11	23.4766	167,601	46	
1NV72LqZAJ...	✗	✗	1	2.16	43	211	59	59	55.6874	135,349	18	

Table 3: Top-10 donation addresses in popular repositories.

Address	# repos	# stars	Avg. age (years)	Donations	
				Amount (dollars)	Volume (txes)
1G8G6tqQ3o...	1	25,406	4.67	3,983	10
1P9BRsmazN...	282	16,904	3.76	0	0
1QDhxQ6Pra...	19	13,095	3.20	9,177	198
1JnC15WwDV...	1	13,082	1.79	0	0
3MDPzjXu2h...	1	11,315	4.55	9	1
13PjuJcfVW...	10	11,057	2.55	0	0
1EMqwwjJr...	3	10,923	5.78	296	1
1DGoNEYAnj...	9	10,371	3.25	29	8
1MDmKC51ve...	1	10,212	5.37	96	6
17NUKd3v7G...	17	10,167	3.59	5	1

repositories, in terms of total number of stars, have received less than the average donation amount of 4,182 dollars, as shown in Table 3. We explore this relationship in more details in §4.3.

Second, donation addresses can receive deposits from channels other than GitHub. In Table 2, we mark an address as tagged if we find a public label describing it with auxiliary information from online services such as WalletExplorer [10], Blockchain.Info [7], and BitcoinWhosWho [6]. We also mark an address as organizational if any of its associated repositories is owned by a GitHub organizational account. For example, the third address in Table 2 is tagged on Blockchain.Info as “Free Software Foundation,” and it turns out to be the foundation’s official donation address [8]. However, none of the associated repositories is owned by the foundation, which suggests that the donations were sent by individuals or organizations other than the users of these repositories. To support this claim, let us consider the top-10 largest deposits to donation addresses, as shown in Table 4. The first deposit, with an amount of over one million dollars, was sent the foundation in early 2018, which is after the repositories were created. However, this deposit transaction is attributed to the anonymous Pineapple Fund, which donated a total of 55 million dollars (5,104 bitcoins) to various charities around the world between late 2017 and early 2018 [16].

Table 4: Top-10 deposits to donation addresses in dollars.

Transaction	Address	Donations	
		Amount (dollars)	Received date (yyyy-mm-dd)
145d80a087...	1PC9aZC4hN...	1,049,380	2018-01-29
72f63ec8b6...	1GbAWBiGyu...	350,550	2017-12-17
a0a1b2dd76...	18xgGTTzZU...	136,377	2017-08-13
6c6af82ea3...	1GbAWBiGyu...	126,390	2017-09-11
e869a0004f...	1GbAWBiGyu...	122,222	2017-09-01
5281ed3ced...	1GbAWBiGyu...	105,248	2017-09-17
a7bcbb803c...	1MEWT2SGbq...	102,852	2017-11-05
c046f593f3...	1GbAWBiGyu...	82,567	2017-09-17
14ee141b36...	33ENWZ9RCY...	77,153	2017-09-21
050a02eadb...	1GbAWBiGyu...	72,039	2017-12-17

Third, donation addresses can receive deposits that are not necessarily donations. As shown in Table 2, nearly all of the addresses are not tagged nor organizational, which means they are likely reused by their owners for various kinds of personal financial transactions. For example, the first address is associated with a repository that is owned by a Github user account from the European Union. The repository hosts PHP code for checking the status of a delegate node in the Shift network [9]. While the repository is unpopular and currently archived (i.e., read-only), the donation address has received six out of the top-10 largest amounts of deposits after the repository was created, constituting over 10% of all donations in dollars, as shown in Table 4. This suggests that such deposits are personal and not actual donations to the repository. Still, we cannot exclude them because there is no evident way to support this claim.

4.2.2 Historical perspective. While it is reasonable to expect that the total amount of donations might increase as Bitcoin’s popularity increases, we found that the exchange rate, which is often described as speculative, seem to have an impact as well. As shown in Figure 2, the total amount of monthly donations in dollars has increased in a relatively small increments until 2017, during which it has increased by orders of magnitude before plummeting down in 2018 onward. Although this change in value resembles the change in bitcoin price in dollars, the resemblance is unclear if we look at the total amount

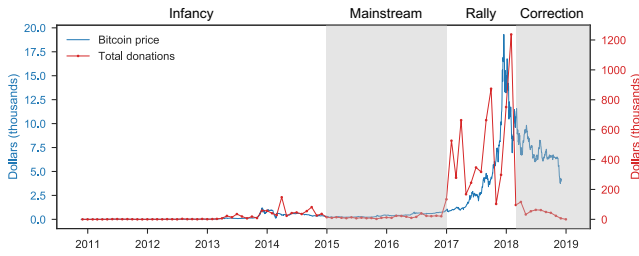


Figure 2: Total amount of donations vs. exchange rate.

Table 5: Top-10 deposits to donation addresses in bitcoins.

Transaction	Address	Donations	
		Amount (bitcoins)	Received date (yyyy-mm-dd)
f9c2fc8da2...	1PC9aZC4hN...	150.0000	2011-05-07
1d36a12fe7...	1PWC7PNHL1...	130.0000	2011-04-30
30bcfe01b...	1abrknajSF...	126.0000	2013-07-01
6c6af82ea3...	1QCdQmJYUq...	107.0892	2013-06-09
1a6e430a62...	1PC9aZC4hN...	100.0000	2012-04-06
145d80a087...	1PC9aZC4hN...	91.4500	2018-01-29
80ba4b1429...	1abrknajSF...	76.7460	2012-11-29
0065017b4c...	1abrknajSF...	76.7460	2012-11-29
b9fd76cd68...	1257U991Wq...	75.0000	2017-01-19
abf03f3c25...	1DcZfySDvU...	70.0000	2010-11-27

of donations in bitcoins instead of dollars, as shown in Figure 3. To better understand this historical aspect of bitcoin donations, we next divide the last 10 years into four periods of time:

Pre-2015 infancy. Bitcoin was mostly used by early-adopters and much easier to mine, especially before 2013, when the mining reward was 50 bitcoins. Moreover, the cryptocurrency was difficult to buy or sell, as online exchanges were insecure and under-developed. As such, it is not surprising that people were comfortable giving away large amounts of bitcoins for donations. In fact, seven out of the top-10 largest deposits in bitcoins, which amount to %8.29 of all donations, were received before 2015, as shown in Table 5.

2015–2016 mainstream. Bitcoin started to attract savvy investors, especially after its price reached that of an ounce of gold in early 2014. With a high expectation for higher returns and an increased liquidity from new exchanges, especially in China, people started to buy bitcoins and hold on them as gold. To some extent, this could explain the decrease in the total amount of donations in this period, which is mostly affected by the decrease in large deposits.

2017 rally. Bitcoin surged almost 400% in a year. Thousands of “Bitcoin millionaire” gave rise to a wave of initial coin offerings and sent more than two billion dollars in funding to hundreds of new cryptocurrency projects [5]. This is also accompanied with large amounts of donations from mid 2017 to early 2018, where the top-10 largest deposits fall within this time period and constitute 26.65% of all donations in dollars, as shown in see Table 4.

Post-2017 correction. Bitcoin went through a major price correction, losing more than 80% of its value. People become more careful

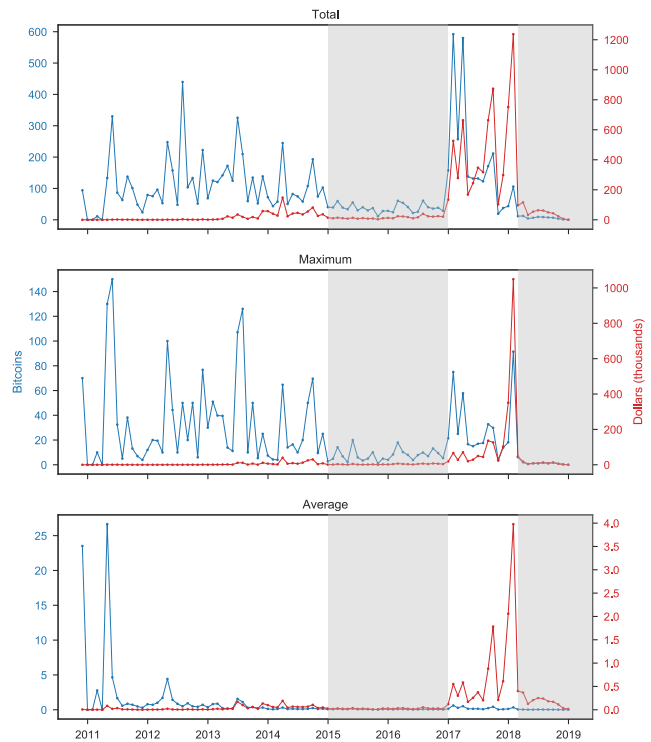


Figure 3: Monthly amount of donations.

with their bitcoins, and the total amount of donations has plummeted to a few thousand dollars a month, without large deposits.

4.3 Popularity and age

As shown in §4.2, some of the addresses that received large amounts of donations were listed in relatively new, unpopular repositories. Our goal is to further explore this relationship and analyze how repository information, its popularity and age in particular, are associated with the amount of donations it receives on listed address. In particular, we perform correlation analysis between five variables namely, the number of watchers, stars, and forks of a repository, its network size, and number of days since the repository was created, and the amount of received donations in bitcoins and dollars.

We note that there are multiple addresses which are listed in one repository, and multiple repositories which list the same address. As repositories and addresses have this many-to-many relationship, we perform a separate analysis for each grouping and aggregation of the data, as follows.

4.3.1 Repository. We grouped the data by each repository identifier, summing the amount of donations from all addresses listed in the repository as the total amount of donations. Figure 4 shows the correlation between repository information, represented as five different variables, and total amount of donations in dollars. We also plot the fit of a linear regression model for each variable with 95% confidence interval. To evaluate pairwise correlation of all variables, including total amount of donations in dollars and bitcoins,

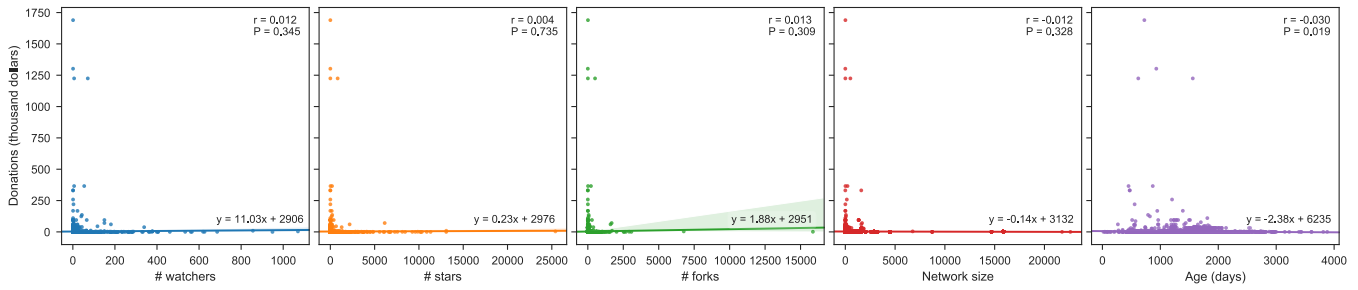


Figure 4: Correlation between repository information and donations.

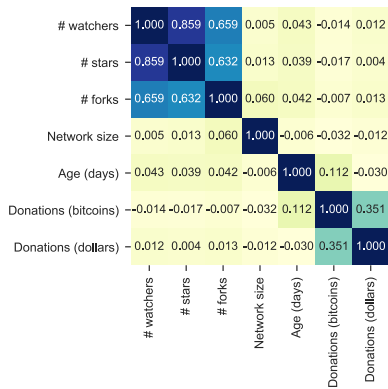


Figure 5: Correlation matrix of repository information.

we also computed the correlation matrix, as shown in Figure 5. In what follows, we discuss the main findings from these two figures.

First, there is a weak positive correlation between number of watchers, stars, and forks and the total amount of donations in dollars, with the strongest correlation being the number of forks with $r=0.013$ and $P=0.309$. However, as shown in Figure 5, there is a negative correlation between these variables and the total amount of donations in bitcoins, which is explained by the deflationary nature of bitcoin price compared to the dollar, even if the two amounts are positively correlated.

Second, from slope values of the fitted models, it is more important to have a larger number of watchers than forks, for instance, in relation to the total amount of donations. In fact, for each new watcher, star, or fork a repository attracts, a total of 11, 0.23, or 1.88 dollars in donations is expected, respectively. To some extent, this suggests that open source software projects could receive more donations if they attract more developers to contribute to their source code, not just users of their software, as software updates typically require forking the repository in order to send the new or modified code with the so called “pull request.”

Third, while there is a weak positive correlation between the age of a repository and its popularity, there is a weak negative correlation between the age of a repository and the amount of donations in dollars. However, this correlation is positive when considering the amount of donations in bitcoins, which is expected, as older repositories did receive more deposits in bitcoins than new

ones, when bitcoin’s price was low and the cryptocurrency was not adopted by mainstream investors.

Fourth and last, the correlations should not be interpreted as statistically significant, as the associated P-values are relatively high, which also means that the amount of donations is considerably influenced by other factors. This can also be explained by the relatively high intercept values of the fitted models, which show that even unpopular repositories that do not have any watchers, stars, or forks, can receive a significant amount of donations on average—at least 2 thousand dollars in this case.

4.3.2 Address. We shift the focus from repositories to addresses, and look at the relationship between address information, represented by aggregated information of repositories where they are listed, and the amount of donations they receive. The rationale behind this grouping is that some popular donation addresses that belong to a specific organization, such as the Free Software Foundation, are used in many repositories that build on top of their software, which could lead to different results.

As such, we grouped the data by each address, aggregating the total (i.e., sum), maximum, and average values per variable, in addition to calculating the number of repositories which list the address. However, as shown in Figure 6, we did not find strong correlation between address information and the amount of donations.

5 DISCUSSION

Bitcoin donations to open source repositories on GitHub are quite limited. On average, a repository receives 4,182 dollars worth of bitcoins in 3.49 years, which is around 100 dollars a month. The speculative and deflationary nature of bitcoin price could explain this disproportionately small amount of donations, as people tend to be discouraged from donating in bitcoins when they have expectations of proportionally higher future price. Another reason might be related to privacy, as once repository owners publicly list donation addresses they become vulnerable to deanonymization attacks, which aim to link their online identities (e.g., GitHub account) to Bitcoin transaction and activities with other users or services [12]. Regardless of the reason, even if it is a behavioral one, it is clear that with such a small amount of donations, repository owners and developers must rely on other sources of income and cannot fully dedicate their time to the hosted projects, unless they are being paid by their employers to work on them. Although open source software is often provided free of charge or not for profit [17], larger amount of donations could surely help in improving its usability,

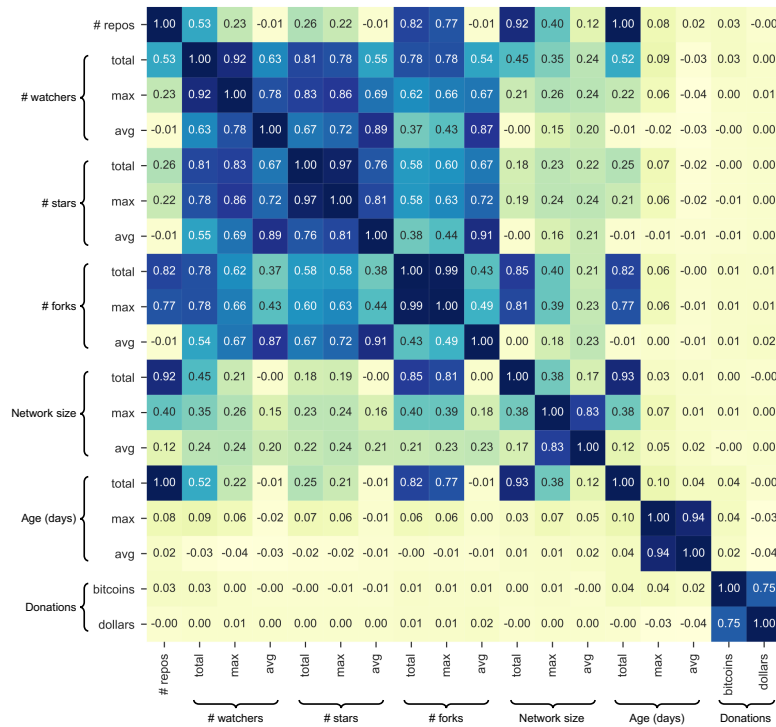


Figure 6: Correlation matrix of address information.

quality, and security by attracting more full-time developers and using better infrastructures.

Repository information, such as popularity and age, do not seem to largely affect the amount of donations a repository is expected to receive. This counter intuitive result could be explained by the fact that some unpopular repositories list personal addresses, which receive deposits for reasons other than donations, or donation addresses of well-known open source organizations, which attract larger amounts of donations from channels other than GitHub.

6 LIMITATIONS

The dataset is limited to licensed open source GitHub repositories, which represent about 10% of all public repositories. While our focus is to analyze Bitcoin donations to open source software, considering all public repositories could lead to new insights about the problem. Moreover, Google BigQuery service hosts only text files on the “HEAD” branch that are less than 1MB in size, without access to file revisions, also called “commit history.” As such, we might have missed extracting some donation addresses because that information exists on a different branch, the address has been added after the snapshot was exported to BigQuery, or because “readme” files are larger than 1MB. While there are other web-based hosting services for open source repositories, such as GitLab [4] and Bitbucket [1], we focus on GitHub as it is the most popular among developers, with more than 96 million repositories.

The analysis is limited to Bitcoin. While our manual address validation shows that developers ask for donations in other cryptocurrency networks, such as Ethereum and Litecoin, Bitcoin remains

the most popular and widely-used network. We do not consider traditional, fee-based payment platforms, such as PayPal for one-off donations and Patreon for monthly donations, because they do not provide access to transaction data. While one could use different data collection methods, such as interviews and surveys, we focus on analyzing donations from evidence-based data sources, such as public blockchains, instead of having to rely on self-reported data.

7 RELATED WORK

Donations are important for the development of open source software. In 2016, LibreOffice, a project of the Document Foundation, reported that they have received more than 200 thousand dollars in donations within a period of three years [30]. The foundation also emphasized that relying on a single sponsor is unnecessary as long as a diverse ecosystem of open source software community exists.

Studying donation economics of open source software is a challenging task. This is especially true for small-size software projects, as donations are typically anonymous and given directly to the people behind the project. Most of the existing research in this domain focuses on interviewing and surveying open source software developers to understand the motivations behind contributing to open source software [14, 23, 24]. To the best of our knowledge, we are the first to analyze the amount and volume of donations from actual financial transactions stored in public blockchains.

DigitalOcean, a cloud infrastructure provider, has recently published as study on developer trends in the open source community [15]. They surveyed 4,300 full-time software developers who work at different companies about whether they use or contribute to

open source software as part of their job. The results show that there is a disconnect between companies' encouragement of open source within their organizations, and their actual investment. While 71% of the respondents reported that their companies expect them to routinely use open source software as part of their day-to-day development work, only 18% of the respondents said their company is a member of an open source-related organization, and only 25% said their company invests more than one thousand dollars every year in donations to open source software.

Ghosh et al. [17] studied the motivations behind contributing to open source software. They surveyed 2,784 developers from the open source software community and found that 4.4% of the respondents would join a software project to make money and 12.3% would stay involved with a project for the money. The authors conclude that developing open source software still resembles rather a hobby than salaried work.

While a large part of the Internet run on open source software, donations to the community, via Bitcoin or otherwise, are small. For example, the OpenSSL library serves 17.5% of all web servers on the Internet, but the OpenSSL Software Foundation receives only two thousand dollars a year in donations [27]. This amount of money is not even enough to pay the developers behind the project, which makes external security audit financially infeasible, leading to infamous bugs like "heartbleed" [29]. Fortunately, after such large-scale security bugs were disclosed, large software companies started the initiative to fund open source software [13].

8 FUTURE WORK

We plan to address the limitation of this study in our future work. In particular, we are currently collecting a larger, more diverse dataset from all public repositories, not just licensed open source ones, from GitHub, GitLab, and Bitbucket. We are also working on enhancing the crawler to fetch the commit history of "readme" files. This will allow us to filter out transactions that were made before an address is listed in a repository, which could be after the repository was created. In addition, we can detect when an address is updated (i.e., changed or removed), and update the analysis accordingly.

We plan to extend the study and consider more cryptocurrency networks, at least Ethereum, in order to get a comparative view of these networks as payment channels for donations. Finally, we plan to conduct a user study with repository owners in order to better understand the use of cryptocurrencies for donations, focusing on usability, security, and privacy issues.

9 CONCLUSION

We conducted a quantitative study analyzing Bitcoin donation to open source software on GitHub. Contrary to the previous research, which is mostly based on self-reported data, our study uses actual financial transactions stored in Bitcoin's public blockchain, which allowed us to calculate the amount and volume of donations over the last 10 years, and explore the relationship between the amount of donations received by a software repository and its hosting-related information, such as its age and popularity on GitHub.

Our results show that Bitcoin donations are unreasonably low, amounting to only 8.3 million dollars worth of bitcoins in the last

10 years, through 44 thousand deposits to six thousand repositories. We did not find strong correlations between the amount of donations and repository popularity or age, where the highest correlation ($r=0.013$, $P=0.309$) was associated with number of forks.

ACKNOWLEDGEMENTS

We would like to thank the people at the Cybersecurity Initiative for Blockchain Research (CIBR) for their help and feedback.³

REFERENCES

- [1] 2018. *Bitbucket - The Git solution for professional teams*. Retrieved 20/12/2018 from <https://bitbucket.org>
- [2] 2018. *CoinMarketCap - Cryptocurrency Market Capitalizations*. Retrieved 23/12/2018 from <https://coinmarketcap.com>
- [3] 2018. *GitHub - The world's leading software development platform*. Retrieved 20/12/2018 from <https://github.com>
- [4] 2018. *GitLab - The first single application for the entire DevOps lifecycle*. Retrieved 20/12/2018 from <https://gitlab.com>
- [5] 2019. *2017 Will Be Remembered As The Year Of Bitcoin*. Retrieved 12/01/2019 from <https://bit.ly/2RPZ0iG>
- [6] 2019. *BitcoinWhosWho - Bitcoin address lookup, checker, and alerts*. Retrieved 12/01/2019 from <https://bitcoinwhoswho.com/>
- [7] 2019. *BlockchainInfo - Search the Blockchain*. Retrieved 12/01/2019 from <https://blockchain.info/>
- [8] 2019. *Free Software Foundations - Ways to Donate*. Retrieved 12/01/2019 from <https://www.fsf.org/about/ways-to-donate/>
- [9] 2019. *The Shift project*. Retrieved 01/02/2019 from <https://www.shiftnrg.org/>
- [10] 2019. *WalletExplorer - Smart Bitcoin block explorer*. Retrieved 12/01/2019 from <https://www.walletexplorer.com/>
- [11] 2019. *Yahoo Finance - Business Finance, Stock Market, Quotes, News*. Retrieved 12/01/2019 from <https://finance.yahoo.com/>
- [12] Yazan Boshmaf, Husam Al Jawaheri, and Mashael Al Sabah. 2019. BlockTag: design and applications of a tagging system for blockchain analysis. In *IPSec Security'19*. Springer, 299–313.
- [13] Jon Brodtkin. 2014. *Tech giants, chastened by Heartbleed, finally agree to fund OpenSSL*. Retrieved 01/02/2019 from <https://goo.gl/QXwPHu>
- [14] Paul A David, Andrew Waterman, and Seema Arora. 2003. FLOSS-US the free/libre/open source software survey for 2003. *Stanford Institute for Economic Policy Research* (2003). <https://goo.gl/JdU2Cp>
- [15] DigitalOcean. 2018. *A Seasonal Report on Developer Trends in the Cloud: Open Source Edition*. <https://goo.gl/ryie9B>
- [16] Free Software Foundation. 2019. *Free Software Foundation receives \$1 million donation from Pineapple Fund*. Retrieved 01/02/2019 from <https://goo.gl/dKvP9K>
- [17] Rishab A Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. 2002. Free/libre and open source software: Survey and Study. Part IV: Survey of developers. <https://goo.gl/HztfwX>
- [18] GitHub. 2018. *About READMEs - User Documentation*. Retrieved 23/12/2018 from <https://help.github.com/articles/about-readmes/>
- [19] GitHub. 2018. *GitHub Activity Data*. Retrieved 05/12/2018 from <https://goo.gl/d4AabN>
- [20] GitHub. 2018. *The State of the Octoverse*. Retrieved 23/12/2018 from <https://octoverse.github.com>
- [21] GitHub: arfon. 2016. *Making open source data more available*. Retrieved 23/12/2018 from <https://goo.gl/jqJvYh>
- [22] Google: Felipe Hoffa. 2016. *GitHub on BigQuery: Analyze all the open source code*. Retrieved 23/12/2018 from <https://goo.gl/pRmwBS>
- [23] A Hars and S OU. 2001. Working for Free?—Motivations of Participating in Open Source Projects, 00 (c), 1–9. In *34th Annual Hawaii International Conference on System Sciences (HICSS-34)*, Hawai. 25–39.
- [24] Guido Hertel, Sven Niedner, and Stefanie Herrmann. 2003. Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research policy* 32, 7 (2003), 1159–1177.
- [25] Harry Kalodner, Steven Goldfeder, Alishah Chator, Malte Möser, and Arvind Narayanan. 2017. *BlockSci: Design and applications of a blockchain analysis platform*. Technical Report. arXiv:cs.CR/1709.02489
- [26] Ambarish Kumar. 2018. *Easily Fund Open Source Projects With These Platforms*. Retrieved 21/12/2018 from <https://goo.gl/7WgiFJ>
- [27] Steve Marquess. 2014. *Of Money, Responsibility, and Pride*. <https://goo.gl/RCdMc8>
- [28] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).
- [29] NetCraft. 2014. *Half a million widely trusted websites vulnerable to Heartbleed bug*. <https://goo.gl/PMYNSH>
- [30] Italo Vignoli. 2016. *200K thanks*. Retrieved 12/1/2019 from <https://goo.gl/g7zyq6>

³For latest research outcomes, please visit <https://qcri.github.io/cibr>