

Incentivized Blockchain-based Social Media Platforms: A Case Study of *Steemit*

Chao Li

School of Computing and Information
University of Pittsburgh
Pittsburgh, USA
chl205@pitt.edu

Balaji Palanisamy

School of Computing and Information
University of Pittsburgh
Pittsburgh, USA
bpalan@pitt.edu

ABSTRACT

Advances in Blockchain and distributed ledger technologies are driving the rise of incentivized social media platforms over Blockchains, where no single entity can take control of the information and users can receive cryptocurrency as rewards for creating or curating high-quality contents. This paper presents an empirical analysis of *Steemit*, a key representative of these emerging platforms, to understand and evaluate the actual level of decentralization and the practical effects of cryptocurrency-driven reward system in these modern social media platforms. Similar to Bitcoin, *Steemit* is operated by a decentralized community, where 21 members are periodically elected to cooperatively operate the platform through the Delegated Proof-of-Stake (DPoS) consensus protocol. Our study performed on 539 million operations performed by 1.12 million *Steemit* users during the period 2016/03 to 2018/08 reveals that the actual level of decentralization in *Steemit* is far lower than the ideal level, indicating that the DPoS consensus protocol may not be a desirable approach for establishing a highly decentralized social media platform. In *Steemit*, users create contents as posts which get curated based on votes from other users. The platform periodically issues cryptocurrency as rewards to creators and curators of popular posts. Although such a reward system is originally driven by the desire to incentivize users to contribute to high-quality contents, our analysis of the underlying cryptocurrency transfer network on the blockchain reveals that more than 16% transfers of cryptocurrency in *Steemit* are sent to curators suspected to be bots and also finds the existence of an underlying supply network for the bots, both suggesting a significant misuse of the current reward system in *Steemit*. Our study is designed to provide insights on the current state of this emerging blockchain-based social media platform including the effectiveness of its design and the operation of the consensus protocols and the reward system.

ACM Reference Format:

Chao Li and Balaji Palanisamy. 2021. Incentivized Blockchain-based Social Media Platforms: A Case Study of *Steemit*. In *Proceedings of ACM Conference on Web Science (WebSci'19)*. ACM, Boston, MA, USA, 10 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebSci'19, July 2019, Boston, MA, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Recent advances in Blockchain and distributed ledger technologies [27] are driving the rise of incentivized social media platforms over Blockchains. Examples of such platforms include *Steemit*¹, *Indorse*², *Sapien*³ and *SocialX*⁴. Unique features of these blockchain-powered social media platforms include decentralization of data generated in the platforms and deep integration of social platforms with the underlying cryptocurrency transfer networks on the blockchain. *Steemit* is the first blockchain-powered social media platform that incentivizes both creator of user-generated content and content curators. It has kept its leading position during the last two and a half years and its native cryptocurrency, *STEEM*, has the highest market capitalization among all cryptocurrencies issued by blockchain-based social networking projects. Its market capital is estimated over 266 million USD in 09/30/2018 [25].

In *Steemit*, users can create and share contents as blog posts. Once posted, a blog can get replied, reposted or voted by other users. Depending on the weight of received votes, posts get ranked and the top ranked posts make them to the front page. All data generated by *Steemit* users are stored in the Steem-blockchain [3]. Similar to other blockchains like Bitcoin [27] and Ethereum [4], data stored in the Steem-blockchain is publicly accessible and it is hard to be manipulated. At the core of *Steemit* are its decentralization and its cryptocurrency-driven reward system used for rewarding the content creators and curators. Instead of operating as a single entity like *Reddit* and *Quora*, the *Steemit* platform is operated by a group of 21 witnesses elected by its shareholders (uses owning *Steemit* shares) through the Delegated Proof of Stake (DPoS) consensus protocol [24]. Unlike traditional social media platforms that typically do not reward their users, *Steemit* issues three types of rewards to its users: (1) producer reward; (2) author reward; (3) curation reward. The producer rewards are issued to the elected witnesses producing blocks. It incentivizes users of *Steemit* to compete for the top-21 witnesses. The author rewards and curation rewards are issued to users creating posts (authors) and users voting for posts (curators) respectively. These incentivize authors to produce posts that attract more votes and curators to vote for posts that have higher potential to be voted by other users. By the end of 2018/08, *Steemit* has issued over 40 million USD worth of rewards to its users [15].

This paper presents an empirical analysis of *Steemit*, a key representative of emerging blockchain-based incentivized social media

¹<https://steemit.com/>

²<https://indorse.io/>

³<https://beta.sapien.network/>

⁴<https://socialx.network/>

platforms. Our study targets two core features of *Steemit*, namely its decentralized operation and its cryptocurrency-driven reward system. By analyzing over 539 million operations performed by 1.12 million users during the period 2016/03 to 2018/08, we aim at obtaining several key insights including the answers for the following set of key questions:

- Do the members of the witness group in the platform have a high update rate or do the same set of users serve in the group? What is the power of big shareholders on the decentralization properties of the platform? Is it possible that a single big shareholder can determine who to join the witness group?
- What are the factors correlated with rewards issued to authors and curators? How does the reward system influence users' behavior? Can incentives be misused by users such as buying votes from bots to promote their posts?

Our analysis reveals interesting details on the decentralized operation in *Steemit* and its reward system. Our study on decentralization in *Steemit* shows that the witness group tends to show a relatively low update rate and its seats may actually be controlled by a few large shareholders. Our analysis also indicates that the majority of top witnesses and top electors form a value-transfer network. These findings together reveals that the actual level of decentralization in *Steemit* is far lower than the ideal level, indicating that DPoS consensus protocol may not be a desirable approach for establishing a highly decentralized social media platform. Our analysis of the reward system in *Steemit* shows that author rewards earned by an author are correlated with a number of factors, including the number of followers, number of created posts and owned *Steemit* shares. However, our analysis of the *transfer* and *vote* operations show that more than 16% transfers in the dataset are sent to curators suspected to be bots. A deeper analysis also reveals the existence of an underlying supply network for the bots, where big shareholders delegate their *Steemit* shares to bots to earn profit. These results together suggests that the current cryptocurrency-driven reward system in *Steemit* is under substantial misuse that deviates from the original intended goal of rewarding high-quality contents. We point out that the current consensus protocols and reward systems can hardly achieve their design targets and identify the key reasons and suggest potential solutions. We believe that the findings in this paper can facilitate the improvement of existing blockchain-based social media platforms and the design of future blockchain-powered websites.

2 BACKGROUND

In this section, we introduce the *Steemit* social media platform that runs over the Steem-blockchain[3]. We present the key use cases of *Steemit* and discuss how *Steemit* leverages the underlying blockchain to function as a decentralized social site that incentivizes users with cryptocurrency-based rewards.

Steemit. Users of *Steemit* can create and share contents as blog posts. A blog post can get replied, reposted or voted by other users. Based on the weights of received votes, posts get ranked and the top ranked posts make them to the front page.

Steem-blockchain. *Steemit* uses the Steem-blockchain[3] to store the underlying data of the platform as a chain of blocks. Every three

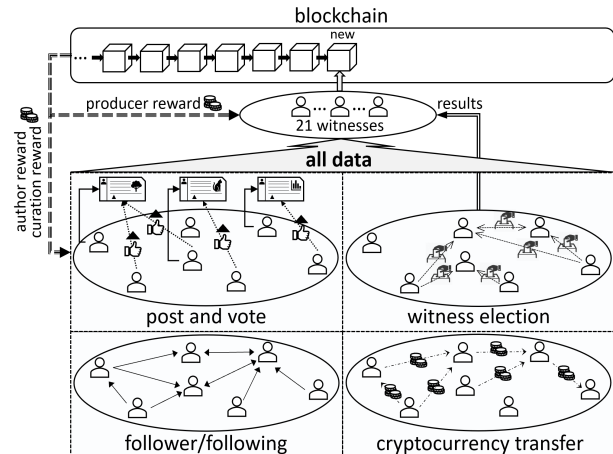


Figure 1: Steem blockchain overview

seconds, a new block is produced, which includes all confirmed operations performed by users during the last three seconds. *Steemit* allows its users to perform more than thirty different type of operations. In Figure 1, we display four categories of operations that are most relevant to the analysis presented in this paper. While post/vote and follower/following are common features offered by social sites (e.g., Reddit [31] and Quora [35]), operations such as witness election and cryptocurrency transfer are features specific to blockchains.

Witness election and DPoS. Witnesses in *Steemit* are producers of blocks, who continuously collect data from the entire network, bundle data into blocks and append the blocks to the Steem-blockchain. The role of witnesses in *Steemit* is similar to that of miners in Bitcoin [27]. In Bitcoin, miners keep solving Proof-of-Work (PoW) problems and winners have the right to produce blocks. With PoW, Bitcoin achieves a maximum throughput of 7 transactions/sec [7]. However, transaction rates of typical mainstream social sites are substantially higher. For example, Twitter has an average throughput of more than 5000 tweets/sec [17]. Hence, *Steemit* adopts the Delegated Proof of Stake (DPoS) [24] consensus protocol to increase the speed and scalability of the platform without compromising the decentralized reward system of the blockchain. In DPoS systems, users vote to elect a number of witnesses as their delegates. In *Steemit*, each user can vote for at most 30 witnesses. The top-20 elected witnesses and a seat randomly assigned out of the top-20 witnesses produce the blocks. With DPoS, consensus only needs to be reached among the 21-member witness group, rather than the entire blockchain network like Bitcoin. This significantly improves the system throughput.

Cryptocurrency - shares and rewards. In *Steemit*, each vote casting a post or electing a witness is associated with a weight that is proportional to the shares of *Steemit* held by the voter. Like most blockchains, the Steem-blockchain issues its native cryptocurrencies called *STEEM* and Steem Dollars (*SBD*). To receive shares of *Steemit*, a user needs to 'lock' *STEEM/SBD* in *Steemit* to receive Steem Power (*SP*) at the rate of $1 \text{ STEEM} = 1 \text{ SP}$ and each *SP* is assigned about 2000 vested shares (*VESTS*) of *Steemit*. A user may withdraw invested *STEEM/SBD* at any time, but the claimed fund will be automatically split into thirteen equal portions to be

withdrawn in the next thirteen subsequent weeks. For example, in day 1, Alice may invest 13 *STEEM* to *Steemit* that makes her vote obtain a weight of 13 *SP* (about 26000 *VESTS*). Later, in day 8, Alice may decide to withdraw her 13 invested *STEEM*. Here, instead of seeing her 13 *STEEM* in wallet immediately, her *STEEM* balance will increase by 1 *STEEM* each week from day 8 and during that period, her *SP* will decrease by 1 *SP* every week.

Through Steem-blockchain, *Steemit* issues three types of rewards to its users: (1) producer reward; (2) author reward and (3) curation reward. The amount of producer reward is about 0.2 *STEEM* per block in 2018/08, meaning that a witness producing blocks for a whole day can earn about 14,400 *STEEM*. Each day, the Steem-blockchain issues a number of *STEEM* (about 53,800 *STEEM* per day in 2018/08) to form the post reward pool and posts compete against each other to divide up the reward pool based on the total weight of votes received within seven days from the post creation time. Here, 75% of reward received by a post goes to the post author and the rest is shared by the post curators based on their vote weight.

In the rest of this paper, for ease of exposition and comparison, we transfer all values of *STEEM/SBD/SP/VESTS* to US dollars (*USD* denoted by \$), based on their median transfer rate from 2018/01 to 2018/08, which is $\$1 = 1 \text{ SBD} \approx 0.4 \text{ STEEM} = 0.4 \text{ SP} \approx 800 \text{ VESTS}$ [14].

3 DATASET AND PRELIMINARY RESULTS

In this section, we describe our data collection methodology and present some preliminary results including the growth of *Steemit* over time and the usage of operations in the platform.

3.1 Data collection

The Steem-blockchain offers an Interactive Application Programming Interface (API) for developers and researchers to collect and parse the blockchain data [13]. From block 1 (created at 2016/03/24 16:05:00) to block 25,563,499 (created at 2018/09/01 00:00:00), we collected 539,817,204 operations performed by 1,120,166 users of *Steemit* during the period 2016/03 to 2018/08. The collected data also includes 121,619,828 virtual operations performed in the *Steemit* platform, such as issuing rewards to witnesses, authors and curators, during the same time period. In the data collected, we recognized 36 types of operations performed by users and 11 types of virtual operations. In table 1, we summarize the key operations (OP) and virtual operations (VO) focused in this paper. We have recently released a dataset for *Steemit* named *SteemOps* at:

<https://github.com/archer1cl1c/SteemOps>

3.2 Preliminary results

Growth of *Steemit*. We first investigate the growth of *Steemit* over time and find that the platform growth is highly impacted by the cryptocurrency market. In Figure 2, we plot the numbers of newly registered users and newly performed operations per month and the changes of *Bitcoin* and *STEEM* price during the period 2016/03 to 2018/08. Apart from the initial boost in the first month, the platform witnessed three times of robust growth, which happened during 2016/05 to 2016/07, 2017/04 to 2017/06 and 2017/11 to 2018/01, respectively. We can observe a strong correlation between the monthly increment of users and the changes in the *STEEM* price.

| OP (social) | Description |
|-------------------------|---|
| comment | users create posts, reply to posts or replies |
| vote | users vote for posts |
| custom_json | users follow other users, repost a blog |
| OP (witness-election) | Description |
| witness_update | users join the witness pool to be elected, witnesses in pool update their information |
| witness_vote | users vote for witnesses by themselves |
| witness_proxy | users cast votes to the same witnesses voted by another user by setting that user as their election proxy |
| OP (value-transfer) | Description |
| transfer | users transfer <i>STEEM/SBD</i> to other users |
| transfer_to_vesting | users transfer <i>STEEM/SBD</i> to <i>VESTS</i> |
| delegate_vesting_shares | users delegate <i>VESTS</i> to other users |
| withdraw_vesting | users transfer <i>VESTS</i> to <i>STEEM</i> |
| VO (reward) | Description |
| producer_reward | platform issues rewards to producers |
| author_reward | platform issues rewards to authors |
| curation_reward | platform issues rewards to curators |

Table 1: Summary of operations and virtual operations

This is due to the fact that more *Steemit* users investing in trading may drive up the *STEEM* price while higher *STEEM* price may in turn attract more people joining *Steemit*. Next, in cryptocurrency market, the changes of *Bitcoin* price are usually seen as the most important market signals. During all the three rising periods, we see the surge in *Bitcoin* price, which illustrates the high influence of the cryptocurrency market on the growth of *Steemit* and also suggests that most *Steemit* users may have a background understanding of cryptocurrency and blockchain. Finally, by comparing the user curve and the operation curve, we find that even though the user growth rate drops after all the three rising periods, the operation growth rate keeps maintaining stability after boosting, which may reflect that most users who joined *Steemit* during the rising periods remain active after the end of the rising periods.

Usage of operations. As discussed in section 3.1, users in *Steemit* may perform a variety of operations and these operations are recorded by the Steem-blockchain. By investigating the usage of these operations, we aim to answer the following questions: (1) which categories of operations are more frequently performed by users?, (2) do users use *Steemit* more like a social media platform or like a cryptocurrency wallet? In Figure 3 and Figure 4, we plot the numbers of social/witness-election operations and value-transfer operations performed in different months, respectively. The functionality of these operations is described in Table 1. Among the three categories of operations, the social operations show the highest utilization rate, which indicates that users are using more social functions offered by *Steemit* than transfer functions. Among the three social operations, the *vote* operation is the most frequently used one. In 2018/08, more than 21 million votes were cast to posts. Unlike other voting-based social media platforms such as *Reddit*, in *Steemit*, votes cast by users owning *Steemit* shares have real monetary value, which may incentivize *Steemit* users to keep voting for posts with some frequency to avoid wasting their voting power. We will discuss more about voting and rewarding mechanisms in

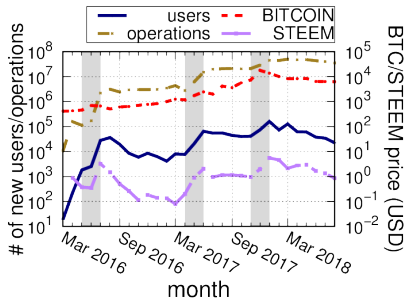


Figure 2: New users/operations per month (2016/03 to 2018/08)

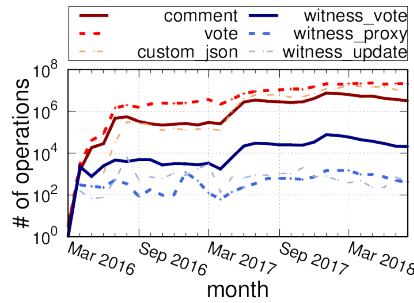


Figure 3: New social/witness-election operations per month (2016/03 to 2018/08)

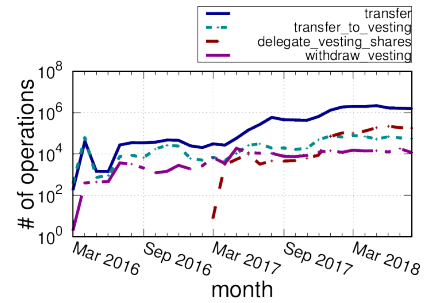


Figure 4: New value-transfer operations per month (2016/03 to 2018/08)

section 5. Among the four value-transfer operations, users perform the *transfer* operation more frequently. Since the *transfer* operation is the only operation among the four that is not associated with *VESTS*, namely shares of *Steemit*, the fact may reflect that more trading behaviors are happening in *Steemit* than investing behaviors. Finally, the number of performed witness-election operations are relatively small, compared with the other two categories. Each month, we see only thousands of users setting or updating their witness votes through the *witness_vote* and *witness_proxy* operations and only hundreds of users joining the witness pool or updating their witness information through the *witness_update* operation. This reflects a relatively low participation rate in witness election regarding both elector and electee and could impact the level of decentralization in the platform. We discuss witness election in more detail in section 4.

4 DECENTRALIZATION IN STEEMIT

As a blockchain-based social media platform, *Steemit* distinguishes itself from traditional social sites through the decentralization brought by the blockchain. In this section, we investigate the actual level of decentralization in *Steemit* by analyzing the witness election process of DPoS in detail.

4.1 Decentralized platform operation

Centralization and decentralization. In traditional social media platforms, such as *Reddit* and *Quora*, a single entity (i.e., *Reddit, Inc.* and *Quora, Inc.*) owns the complete data generated by users and operates the websites. In contrast, *Steemit* not only open sources its front-end, *Condenser* and the back-end *Steem-blockchain* [30], but also makes all its data in the blockchain available for public access [3]. Rather than functioning as a single entity, the *Steemit* platform is operated by a group of 21 witnesses elected by its users. Any user in *Steemit* can run a server, install the *Steem-blockchain* and synchronize the blockchain data to the latest block. Then, by sending a *witness_update* operation to the network, the user can become a witness and have a chance to operate the website and earn producer rewards if he or she can gather enough support from the electors to join the 21-member witness group.

Steemit-2 and STEEM-2. As anyone can copy the code and data of *Steemit*, one may naturally doubt that an adversary, say Bob, may build a ‘fake’ *Steemit* platform, *Steemit-2* that has exactly same the functionality and historical data as *Steemit*. To distinguish *Steemit-2*

from *Steemit*, we name the cryptocurrency issued by *Steemit-2* as *STEEM-2* (and also *SBD-2*). Here, a natural question that arises is what makes people believe that *Steemit*, rather than *Steemit-2*, is the ‘real’ one? In the decentralized network, the opinion of ‘which one is real’ is determined by the consensus of *Steemit* users or, to be more precise, the shareholders. With the DPoS consensus protocol, each block storing data of *Steemit* is signed by a top witness elected by the shareholders, which may represent the consensus among the shareholders. Therefore, unless most of shareholders switch to *Steemit-2*, the new blocks generated in *Steemit-2* will be signed by witnesses elected by a few shareholders and will not be recognized by the entire community.

Factors affecting decentralization of Steemit. In our work, we study the characteristics of decentralization in *Steemit* by analyzing the witness election process. In general, we consider *Steemit* to have an ideal level of decentralization if members of the 21-member witness group are frequently updated and if these members all have different interests. We also consider *Steemit* to have a relatively high decentralization closer to the ideal level if it allows more people to join the 21-member witness group, if the power of big shareholders is not decisive and if the election is not heavily correlated with value-transfer operations. We investigate these aspects in the following subsection.

4.2 Analyzing witness election

Update of the 21-member witness group. To investigate the update of the 21-member witness group, we extract the producer of each block from block 1 to block 25,563,499 and plot the results as a heatmap in Figure 5. We first compute the number of blocks produced by each witness and sort the witnesses based on their produced blocks in total. For the top-30 witnesses that have the highest number of produced blocks, we plot their attendance rate in the 21-member witness group during the thirty months. During a month that has thirty days, there should be $30 * 24 * 60 * 60 / 3 = 864,000$ blocks generated because blocks are generated every three seconds. For every 21 blocks (63 seconds), the 21 elected witnesses are shuffled to determine their order for generating the next 21 blocks. Therefore, if a witness has a 100% attendance rate in the elected group, it can at most produce $864000 / 21 \approx 41,142$ blocks in a thirty-day month. In Figure 5, we find that most of the top-30 witnesses, once entering the 21-member witness group, maintained a high attendance rate closer to 100% for a long time. From month 12

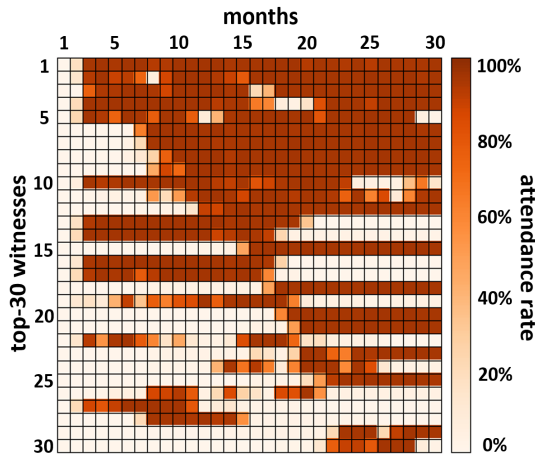


Figure 5: Heatmap of top-30 witnesses' attendance rate in the 21-member witness group during 30 months from 2016/03 to 2018/08

to month 30, namely one and a half year, the top-12 witnesses firmly held at least 10 seats, nearly half of positions in the 21-member witness group. From month 23 to month 30, 17 seats were held by 18 witnesses. From witness 13 to witness 30, we can observe a transition period, namely month 15 to month 20, during which the places of nine old witnesses were gradually taken by nine new witnesses. Overall, the 21-member witness group tends to show a relatively low update rate. The majority of seats were firmly controlled by a small group of witnesses. We do observe some switch of seats but that happened only in a low frequency.

Power of big shareholders. Next, we investigate the influence of big shareholders in the witness election. As described in Table 1, a user has two ways to vote for witnesses. The first option is to perform *witness_vote* operations to directly vote for at most 30 witnesses. The second option is to perform a *witness_proxy* operation to set another user as an election proxy. For example, Alice may set Bob to be her proxy. Then, if both Alice and Bob own \$100 worth of shares, any vote cast by Bob will be associated with a weight of \$200 worth of shares. Once Alice deletes the proxy, the weight of Bob's votes will reduce to \$100 worth of shares immediately. In Figure 6, we plot the stacked bar chart representing a snapshot of weighted votes received by the top-60 witnesses, who have produced the highest amounts of blocks, at block 25,563,499. The figure shows the distribution of votes cast by the top-29 electors whose votes have the highest weight, either brought by their own shares or shares belonging to users setting these electors as proxy. The sum of weighted votes cast by all other electors outside the top-29 is represented as 'sum of rest'. From Figure 6, we see a few top electors have their votes weighted by striking amounts of shares. The top-1 elector has his or her votes weighted by \$19,800,000 worth of shares. A deeper investigation regarding the top-1 elector shows that all the shares affecting his or her weight are not directly owned by this user, but owned by another user, who set the top-1 elector as proxy. The runner-up elector, which is the account belonging to the main exchange used by *Steemit* users, has its votes associated with a weight of \$12,100,000 worth of shares. From the figure, we see all the 27 witnesses voted by the top-1 elector enter the top-50, all

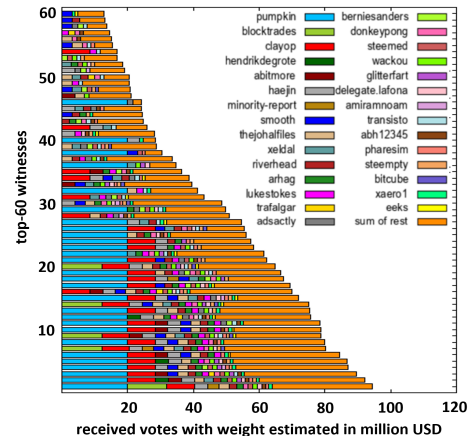


Figure 6: A snapshot of weighted votes received by top-60 witnesses at block 25,563,499. The votes are weighted by shares (estimated in USD) owned by the electors. We show the votes cast by the top-29 electors owning the highest weight and use 'sum of rest' to represent the sum of weighted votes cast by all other electors.

the 18 witnesses receiving votes from both the top-2 electors enter the top-30 and the only witness receiving votes from all the top-3 electors becomes the top-1 witness. We can also observe that 19 out of the top-20 witnesses receive at least two votes from the top-3 electors and 29 out of the top-30 witnesses receive at least one vote from the top-3 electors. As illustrated by the results, the distribution of weight of votes in witness election is heavily skewed, which suggests that the election of 21 witnesses may be significantly impacted by a few big shareholders. This phenomenon may not be a good indication for a decentralized social media platform. In the worst case, if the 21-member witness group is controlled by a single shareholder, the platform will simply function as a centralized model.

Value-transfer operations among election stakeholders. Finally, we investigate the value-transfer operations performed among the top-30 witnesses, top-29 electors and the accounts selecting top-29 electors as proxy. The data is plotted as a directed graph in Figure 7, where edges are colored by their source node color. The edge thickness represents the amount of transferred value from source to target, which is the sum of value transferred through *transfer* and *transfer_to_vesting* operations. Since most *Steemit* users use runner-up elector to trade cryptocurrency, the graph does not show edges connected with that account. Our first observation about this graph is that only two out of top-30 witnesses and three out of top-29 electors never perform value-transfer operations while all other investigated users form a value-transfer network, which has 3.34 average degree, 0.21 average clustering coefficient and 3.96 average path length. In the network, we find a cluster of users who select top-29 electors as proxy. After manually checking the profiles of these users, we find that this cluster represents a community of Korean users, which is connected to the rest of the network mainly through several leaders of the Korean community. Overall, what we observed from the value-transfer operations suggests that the majority of the investigated election stakeholders have economic interactions, which may not be a good indication of a perfectly

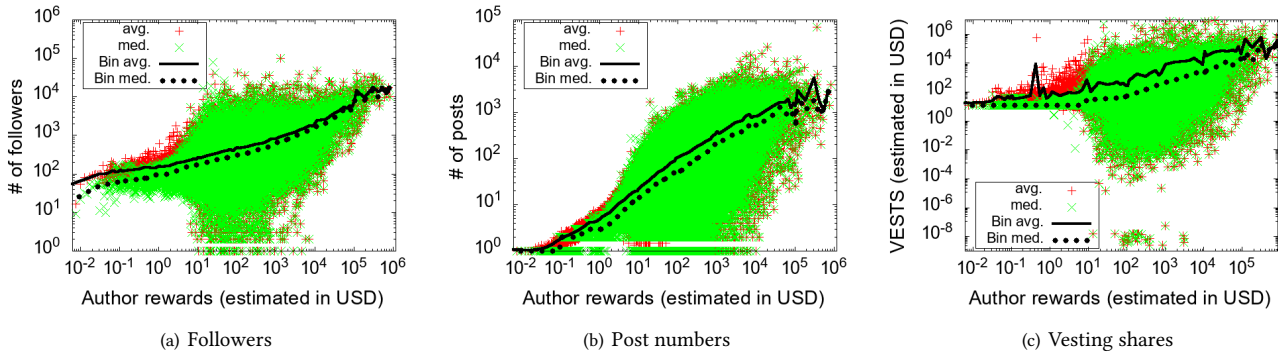


Figure 8: Scatter plot investigating correlation between followers (post numbers, vesting shares) and author rewards

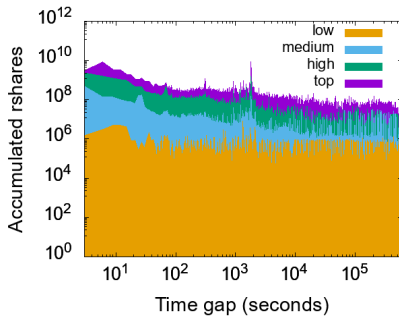


Figure 9: Distribution of accumulated $rshares$ over time after post creation time

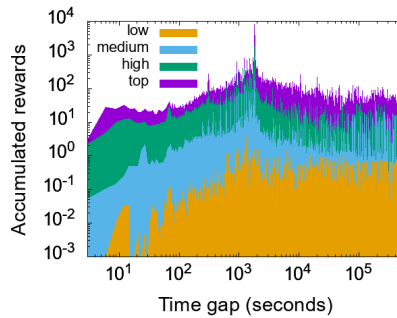


Figure 10: Distribution of accumulated curation rewards over time

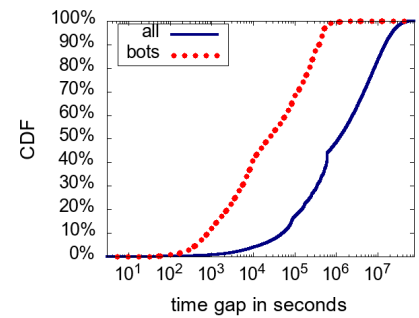


Figure 11: CDF investigating the time gap between post creation time and bot voting time

As can be seen, cr is computed from 25% of pr , but is affected by two factors. The first factor is the ratio between $\sqrt{rs_b + rs} - \sqrt{rs_b}$ and $\sqrt{rs_T}$, where rs_b denotes the $rshares$ accumulated by the post before this user's vote and rs_T refers to the total accumulated $rshares$ by the post during the 7-day time window. This is a very interesting factor as it suggests that curators who want to earn higher cr should vote for posts as early as possible to make rs_b smaller. It also suggests that such curators vote for posts that have higher probability to accumulate higher rs_T by the end of 7-day time window. However, the second factor forces curators to reconsider the early vote strategy. It is the minimum of $\frac{td}{30}$ and 1, where td denotes the time difference in minutes between the post creation time and voting time. For example, if Alice votes for a post one minute after the post creation time, she will only receive $\frac{1}{30}$ of the curation rewards assigned to her while the rest $\frac{29}{30}$ is again issued to the post author. Instead, if Alice casts her vote 30 minutes after the post creation time, she can completely earn the assigned curation rewards. Because of the second factor, a post author usually receives more than 75% of the post reward. In summary, the two factors in the curation reward equation make it hard to determine the best strategy of the voting time in the reward earning game. However, it is clear that voting for posts that are likely to be voted by more users in the future is certainly helpful.

5.2 Factors correlated with rewards

After theoretically analyzing the reward system, we now investigate the blockchain data to learn the factors correlated with author rewards and curation rewards.

Author rewards. Regarding author rewards, we investigate three factors: (1) number of followers owned by an author; (2) number of posts created by an author; (3) amount of vesting shares ($VESTS$) owned by an author. In order to gauge the correlation between these factors and author rewards, for each factor, we plot in Figure 8 the average and median of values of the factor (y) against author rewards estimated in USD (x). Following the methodology used by Kwak *et al.* [23], we also bin the value of author rewards in log scale and plot the average and median per bin in lines. In Figure 8(a), we see the line of average is always above the line of median, indicating that some users with followers more than average do not receive expected degree of support from their followers regarding their posts. The gap between the two line is larger for users receiving lower author rewards, but becomes relatively small for the top authors earning rewards more than \$1000. The median number of followers grows steadily without showing a flat period or a surge. Next, in Figure 8(b), we investigate correlation between number of posts created by an author and rewards received by an author. Again, we find that the line of average keeps itself above the line of median, indicating that some users post more blogs than average but do not receive more rewards as they may expect. This may suggest that some authors slow down their posting speed while improving

the quality of their posts. Until \$30,000 rewards, the median number of posts shows a relatively stable growth. After that, it experiences a fluctuation among the top authors. Finally, in Figure 8(c), we investigate correlation between VESTS owned by an author and rewards received by an author. From the observation that the line of median lies nearly always below the line of average, we may infer that some users own higher shares of *Steemit* than average but fail to increase their author rewards to the same scale. By observing the line of median, we find that the median number of VESTS stays relatively flat at \$12.5 before author rewards reach \$10. The \$12.5 worth of VESTS is equal to 5 *STEEM*, which is the minimum line that *Steemit* suggests that its user transfer to *VESTS* for the purpose of maintaining the user's account in a normal state. From \$10 worth of author rewards, the median points start to disperse. It is interesting to find that many authors earning rewards more than \$10 maintain very low *VESTS*, even lower than the suggested minimum line. However, beyond \$10 worth of author rewards, the line of median still shows a positive trend.

Overall, our results illustrate a positive correlation between all the three factors and author rewards. Our results do not imply causation, but they suggest that users, both authors and curators, consider the three factors as indicators of the potential popularity of posts.

Curation rewards. As we have analyzed in section 5.1, the curation reward assigned to a single curator is affected by two factors. The first factor suggests that reward-driven curators cast their votes early and vote for posts with high potential for popularity. However, the second factor penalizes the early votes. Regarding the potential popularity of posts, a reward-driven curator may leverage the number of followers/posts/VESTS of post authors as indicators. However, due to the conflict suggestions from the two factors in the curation reward computation equation, it is hard to determine the best voting time for maximizing curation rewards. Therefore, we decide to investigate what are the choices made by users in the data set. To observe decisions made by different levels of shareholders, we divide users into four groups based on their owned *VESTS*, namely low ($< 10^4$), medium ($10^4 - 10^6$), high ($10^6 - 10^8$) and top ($> 10^8$). In Figure 9, we plot the distribution of the accumulated *rshares* that have been contributed by all votes to all posts every second after the post creation time. Since *rshares* is the most important factor of post rewards and also the key value of votes, this figure helps revealing users' strategies on selecting voting time. As can be seen from results, there are two peaks that happened. The first peak is at the sixth second, which indicates that many curators choose to vote for a post after six seconds (two blocks) of the post creation time. This may reflect that many users are still using the early vote strategy. The second peak is more interesting. It happens at the 1803 second, namely the first block after the 30-minute time period penalized by the second factor of the curation reward computation equation. This reflects that some users do understand the punishment mechanism and are deliberately avoiding from being penalized. We can also observe that the low-level shareholders tend to not pay special attention to the voting strategies while all the other three groups of shareholders tend to contribute more to the two peaks and are more likely to vote earlier besides the two peaks. The results of the distribution of accumulated curation rewards along time are shown in Figure 10, which show a clear effect of the

punishment mechanism on the early votes. The results suggest that curators who want to earn more rewards vote around the time of the second peak, namely the end of the 30-minute penalty period.

5.3 Misuse of reward system

The reward system in *Steemit* may be a good way to reward contributors to the platform, but it may also be misused by some users in ways that deviate from the original intended goal in *Steemit*, such as buying votes from bots to promote some meaningless posts for the purpose of earning profit. In this subsection, we aim at understanding to what extent such behaviors are performed in *Steemit*.

Temporal correlation. We study this topic by investigating the temporal correlation between *transfer* operations and *vote* operations. Concretely, if there is a suspicious vote cast by a curator to an author through a *vote* operation, there should also be a suspicious fund transferred to the curator from the author through a *transfer* operation that happens before the *vote* operation is performed by the curator. Specifically, we consider a *transfer* operation is suspicious if the 'memo' area (allowing sender to leave a message) of the *transfer* operation only contains a link pointing to a recent post created by the sender. If in addition, the recipient of the *transfer* operation, after receiving the fund from the sender, votes for the post matching that link within the 7-days time window after the post creation time, we consider it as a suspicious trade between a post author and a voting bot.

After parsing the 18,172,530 *transfer* operations in blockchain, we find that 5,031,737 (27.69%) of them only contain a single post link in the 'memo' area. By further investigating the *vote* operation, we find that 2,939,051 (16.17%) of the total *transfer* operations are followed by a suspicious *vote* operation voting for the post link within the 7-days time window, which are performed by the recipients of the funds. Among all the fund recipients, the top one has performed such trade for 113,068 times and earned nearly 1 million USD.

Voting time of suspicious curators. To further investigate the features of the curators suspected to be bots, in Figure 11, we plot the cumulative distribution function (CDF) of the time gap between the post creation time and voting time. The solid line refers to the CDF of votes cast by all curators who ever received any amount of fund from the authors of voted posts at any time point in the past. The dotted line describes the CDF of votes cast by curators suspected to be bots. As can be seen, the curators suspected to be bots cast their votes much earlier than the average. The dotted line reaches 98.9% at the 604,800 second, which refers that nearly all their votes are sent within the 7-day time window and they are internationally contributing *rshares* of their votes to the posts.

An example bot network. In Figure 12, we plot an example bot network, which describes the *transfer* operations sent to the curators suspected to be bots from the top-30 post authors, who most frequently had contacted with these suspected curators. The node sizes are weighted by weighted in-degree. As can be seen, the thirty authors contacted a large number of suspected curators because they may need *rshares* from more than one suspected curator to promote their posts. We can also observe that most suspected curators have their account name containing suspected keywords, such as *boost*, *promote*, *whale* and *upvote*.

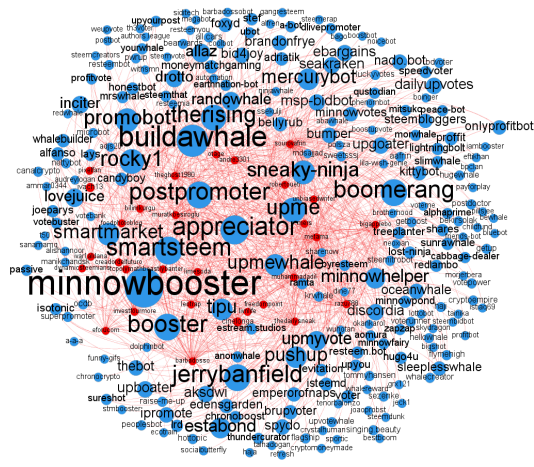


Figure 12: Graph of *transfer* operations sent from selected post authors (red) to the curators suspected to be bots (blue). The graph was plotted using Gephi [2].

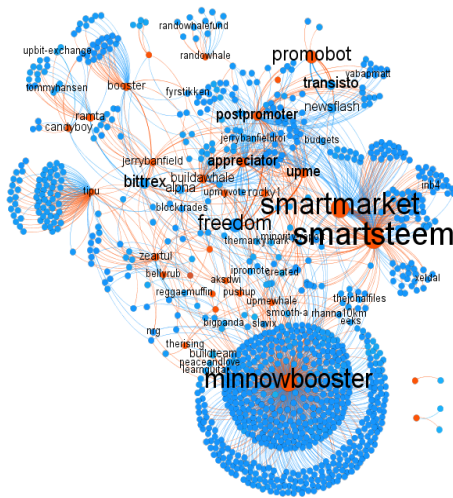


Figure 13: Graph of *delegate_vesting_shares* operations and *transfer* operations among the top-30 curators suspected to be bots (red) and their suspected suppliers (blue). The graph was plotted using Gephi [2].

An example supply network for bots. Finally, we analyze where do these suspected curators get large amounts of *VESTS* to make their votes weighted by higher *rshares*. After a deep analysis, we find that there is an underlying supply network, where some big shareholders delegate their *VESTS* to the suspected curators through *delegate_vesting_shares* operations and the suspected curators periodically pay ‘rent’ to these *VESTS* suppliers through *transfer* operations. In Figure 13, we plot an example supply network for bots, which describes the *delegate_vesting_shares* operations sent to the top-30 suspected curators with highest earnings and the *transfer* operations sent from the top-30 suspected curators with amount higher than \$100. The node size reflects weighted degree and the edge thickness is weighted by the amount of transferred value. From the graph, we find that most of the top-30 suspected curators

own a number of suppliers and a few of them have many. It is interesting to find that the top-1 supplier is the same user who also stands behind the top-1 witness presented in section 4.2. We find that this user has earned more than 1.5 million USD by delegating his or her *VESTS* to the curators suspected to be bots.

5.4 Discussion

In summary, we have discussed the design principle, current use status and underlying misuse of the cryptocurrency-driven reward system of *Steemit* in this section. The rise of social bots and the harm caused by them to the online ecosystems has been widely recognized [9]. Our results in this section reveal that bots have broadly appeared in the emerging incentivized blockchain-based social media networks. In *Steemit*, the monetary value of cryptocurrencies and the deep integration of social operations and value-transfer operations have jointly led to the special supply chain, where bots buy voting powers from big stakeholders and sell the voting powers to users who want to promote posts. Clearly, such behaviours violate the original intention of the reward system and may eventually result in a front page occupied by valueless posts or spams promoted by bots. Moreover, due to the decentralized operation, it is difficult to delete these problem posts appeared at the front page because no single entity has this power. Our discussion in this section also suggests a way of detecting bots through investigating the temporal correlation between *transfer* operations and *vote* operations. We believe this detection approach is effective at the current stage when bots prefer to trade with the native cryptocurrency *STEEM*, which is detectable through the Steem-blockchain. However, smarter bots may hide the temporal correlation between *transfer* operations and *vote* operations by asking their clients to pay them through other cryptocurrencies such as Bitcoin or even privacy-preserving cryptocurrencies such as Zerocash [28]. Then we need to rely on other approaches, such as Deep Neural Networks [22], to detect bots.

6 RELATED WORK

In recent years, due to the rapid growth and consistent popularity, social media platforms have received significant attention from researchers. A large number of research papers have analyzed several popular social media platforms from various perspectives [1, 8, 11, 12, 29, 31, 32, 35, 36]. Tan *et al.* [32] investigated users’ behavior in *Reddit* and found that users continually post in new communities. Singer *et al.* [29] observed a general quality drop of comments made by users during activity sessions. Hessel *et al.* [12] investigated the interactions between highly related communities and found that users engaged in a newer community tend to be more active in their original community. In [11], the authors studied the browsing and voting behavior of *Reddit* users and found that most users do not read the article that they vote on. An extensive survey of recent research on *Reddit* is provided in [26]. Besides *Reddit*, several other social media platforms have also been analyzed by researchers. Wang *et al.* [35] analyzed the *Quora* platform and found that the quality of *Quora*’s knowledge base is mainly contributed by its user heterogeneity and question graphs. Anderson *et al.* [1] investigated the *Stack Overflow* platform and observed significant assortativity in the reputations of co-answeraers, relationships between reputation and answer speed.

Recent advances in blockchain and distributed ledger technologies in terms of scalability [18, 20], efficiency [6, 34] and privacy [5, 21, 28] have empowered blockchains to support various services beyond money transfer, including incentivized blockchain-based social media platforms such as *Steemit*. Recently, this new type of social media platform has drawn some attention from researchers. Thelwall *et al.* [33] analyzed the first posts made by 925,092 *Steemit* users to understand the factors that may drive the post authors higher rewards. Their results suggest that new users of *Steemit* start from a friendly introduction about themselves rather than immediately providing useful content. In a very recent work, Kiayias *et al.* [19] studied the decentralized content curation mechanism from a computational perspective. They defined an abstract model of a post-voting system, along with a particularization inspired by *Steemit*. Through simulation of voting procedure under various conditions, their work identified the conditions under which *Steemit* can successfully curate arbitrary lists of posts and also revealed the fact that selfish participant behavior may hurt curation quality. To the best of our knowledge, the work presented in this paper is the first research that investigates the reward system and decentralization features of incentivized blockchain-based social media platforms through a rigorous empirical analysis of the operations reflected in the underlying blockchain data.

7 CONCLUSION

In this paper, we presented an empirical analysis of *Steemit*, a blockchain-based incentivized social media platform where no single entity can take control of the information and users are rewarded for the contributions they make. We analyzed over 539 million operations performed by 1.12 million users during the period 2016/03 to 2018/08. Our results show interesting details about two core features of *Steemit*, namely its decentralized management and its reward system. Our study on decentralization in *Steemit* shows the actual level of decentralization in *Steemit* is far lower than the ideal level, indicating that DPoS consensus protocol may not be a desirable approach for establishing a highly decentralized social media platform. Our analysis of the reward system reveals the fact that more than 16% transfers of cryptocurrency in *Steemit* are sent to curators suspected to be bots and also finds the existence of an underlying supply network for the bots, both suggesting that the current cryptocurrency-driven reward system in *Steemit* is under severe misuse that deviates from the original intended goal of rewarding high-quality contents. Overall, we believe that the results in this paper provide insights on the current state of the emerging blockchain-based social media platforms including the effectiveness of the design and the operation of the consensus protocols and the reward system.

REFERENCE

- [1] Ashton Anderson et al. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD*, pages 850–858. ACM, 2012.
- [2] Mathieu Bastian et al. Gephi: an open source software for exploring and manipulating networks. *Icwsm*, 8(2009):361–362, 2009.
- [3] Steem blockchain [Internet]. Available from. <https://developers.steem.io/>. Accessed Oct. 2018.
- [4] Vitalik Buterin et al. A next-generation smart contract and decentralized application platform. *white paper*, 2014.
- [5] Jan Camenisch, Manu Drijvers, and Maria Dubovitskaya. Practical uc-secure delegatable credentials with attributes and their application to blockchain. In *Proceedings of the 2017 CCS*, pages 683–699. ACM, 2017.
- [6] Melissa Chase and Sarah Meiklejohn. Transparency overlays and applications. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 168–179. ACM, 2016.
- [7] Kyle Croman et al. On scaling decentralized blockchains. In *International Conference on Financial Cryptography and Data Security*, pages 106–125, 2016.
- [8] Munmun De Choudhury and Sushovan De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*, 2014.
- [9] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Communications of the ACM*, 59(7):96–104, 2016.
- [10] Yossi Gilad, Rotem Hemo, Silvio Micali, Georgios Vlachos, and Nickolai Zeldovich. Algorand: Scaling byzantine agreements for cryptocurrencies. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 51–68. ACM, 2017.
- [11] Maria Glenski, Corey Pennycuff, and Tim Weninger. Consumers and curators: Browsing and voting patterns on reddit. *IEEE Transactions on Computational Social Systems*, 4(4):196–206, 2017.
- [12] Jack Hessel et al. Science, askscience, and badscience: On the coexistence of highly related communities. In *ICWSM*, pages 171–180, 2016.
- [13] Interactive Steem API [Internet]. Available from. <https://steem.eesteem.ws/>. Accessed Oct. 2018.
- [14] STEEM Price [Internet]. Available from. <https://coinmarketcap.com/currencies/steem/>. Accessed Oct. 2018.
- [15] Steem Total Rewards [Internet]. Available from. <https://steem.io/>. Accessed Oct. 2018.
- [16] Steem Whitepaper [Internet]. Available from. <https://steem.io/steem-whitepaper.pdf>. Accessed Oct. 2018.
- [17] Twitter Usage Statistics [Internet]. Available from. <http://www.internetlivestats.com/twitter-statistics/>. Accessed Oct. 2018.
- [18] Harry Kalodner et al. Arbitrum: Scalable, private smart contracts. In *Proceedings of the 27th USENIX Conference on Security Symposium*, pages 1353–1370. USENIX Association, 2018.
- [19] Aggelos Kiayias, Benjamin Livshits, Andrés Monteoliva Mosteiro, and Orfeas Stefanos Thyfronitis Litos. A puff of steem: Security analysis of decentralized content curation. *arXiv preprint arXiv:1810.01719*, 2018.
- [20] Eleftherios Kokoris-Kogias et al. Omniledger: A secure, scale-out, decentralized ledger via sharding. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 583–598. IEEE, 2018.
- [21] Ahmed Kosba et al. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In *2016 IEEE symposium on security and privacy (SP)*, pages 839–858. IEEE, 2016.
- [22] Sneha Kudugunta and Emilio Ferrara. Deep neural networks for bot detection. *Information Sciences*, 467:312–322, 2018.
- [23] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
- [24] Daniel Larimer. Delegated proof-of-stake (dpos). *Bitshare whitepaper*, 2014.
- [25] Steem market cap [Internet]. Available from. <https://coinmarketcap.com/currencies/steem/>. Accessed Oct. 2018.
- [26] Alexey N Medvedev et al. The anatomy of reddit: An overview of academic research. *arXiv preprint arXiv:1810.10881*, 2018.
- [27] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [28] Eli Ben Sasson et al. Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy (SP)*, pages 459–474. IEEE, 2014.
- [29] Philipp Singer, Emilio Ferrara, Farshad Kooti, Markus Strohmaier, and Kristina Lerman. Evidence of online performance deterioration in user sessions on reddit. *PLoS one*, 11(8):e0161636, 2016.
- [30] Steemit source code [Internet]. Available from. <https://github.com/steemit>. Accessed Oct. 2018.
- [31] Greg Stoddard. Popularity and quality in social news aggregators: A study of reddit and hacker news. In *Proceedings of the 24th WWW*, pages 815–818. ACM, 2015.
- [32] Chenhao Tan and Lillian Lee. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the 24th WWW*, pages 1056–1066, 2015.
- [33] Mike Thelwall. Can social news websites pay for content and curation? the steemit cryptocurrency model. *Journal of Information Science*, page 0165551517748290, 2017.
- [34] Alin Tomescu and Srinivas Devadas. Catena: Efficient non-equivocation via bitcoin. In *2017 38th IEEE Symposium on Security and Privacy (SP)*, pages 393–409. IEEE, 2017.
- [35] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y Zhao. Wisdom in the social crowd: an analysis of quora. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1341–1352. ACM, 2013.
- [36] Diyi Yang and Scott Counts. Understanding self-narration of personally experienced racism on reddit. In *ICWSM*, pages 704–707, 2018.