Can Social News Websites Pay for Content and Curation? The SteemIt Cryptocurrency Model[1]

Mike Thelwall, University of Wolverhampton, UK.

Abstract

SteemIt is a Reddit-like social news site that pays members for posting and curating content. It uses micropayments backed by a tradeable currency, exploiting the Bitcoin cryptocurrency generation model to finance content provision in conjunction with advertising. If successful, this paradigm might change the way in which volunteer-based sites operate. This paper investigates 925,092 new members' first posts for insights into what drives financial success in the site. Initial blog posts on average received $0.01, although the maximum accrued was $20,680.83. Longer, more sentiment-rich or more positive comments with personal information received the greatest financial reward in contrast to more informational or topical content. Thus, there is a clear financial value in starting with a friendly introduction rather than immediately attempting to provide useful content, despite the latter being the ultimate site goal. Follow-up posts also tended to be more successful when more personal, suggesting that interpersonal communication rather than quality content provision has driven the site so far. It remains to be seen whether the model of small typical rewards and the possibility that a post might generate substantially more are enough to incentivise long term participation or a greater focus on informational posts in the long term.

Keywords: News websites; SteemIt; Reddit; Cryptocurrency; social web.

Introduction

The web has hosted many new types of online community, from Instagram to Facebook. Most are free to join, with the site owners generating revenue by selling data and advertising. Participation may be encouraged by the affordances of a site, belief in its goal or the value of contributing to its community (social capital) or by reputation-based non-financial rewards. The SteemIt blog-based social news website goes further by allowing participants to receive micropayments for generating content [1]. Members also receive payments for identifying (upvoting) popular content when it is first posted. Although this is a relatively new website with under two hundred thousand members in July 2017 (the word's 2,187th most popular website according to Alexa in October 2017, with users mainly from the USA: http://www.alexa.com/siteinfo/steemit.com), its innovative model of micropayment for content provision backed by a cryptocurrency may become more widespread in the future, if successful in SteemIt. The design goals and affordances of any new system do not determine how it is used because usage cultures [2] including network interactions (e.g., [3,4]) can emerge that adapt them in unexpected ways. For example, Reddit's features have been argued to accidentally facilitate toxic technocultures [5]. Thus, the usage of SteemIt must be investigated to assess the impact of its micropayment model.

SteemIt uses a cryptocurrency, Steem, that differs from the internal currencies of other online communities, such as Second Life's Linden Dollars [6], by being externally tradable. The core of SteemIt is a set of posts created by users. Authors of blog posts,

---

comments and replies that are upvoted (i.e., voted for) by other users receive Steem Dollars (i.e., Steem, but expressed as the U.S. dollar equivalent) according to a formula involving the number of voters, the "Steem Power" of these voters and the current U.S. dollar/Steem exchange rate [7]. The Steem Dollar rewards can be converted into offline currencies or other cryptocurrencies through currency exchanges. The Steem reward does not cost the voter because it is taken from the SteemIt platform from a central pool [8] and regenerates over time from new centrally-created Steem. Moreover, the first upvoters of content that subsequently receives many upvotes also receive a share of the Steem Dollars as the content "curators" that originally judged it worthy. Users who monitor content and successfully identify future high value posts therefore get paid for this librarian-like function, although only the original poster can categorise the content. Thus, SteemIt is like Reddit except the two key functions – finding and creating content – may be financially rewarded. At the technical level SteemIt posts and user information are stored in a decentralised blockchain model. There are two other ways of receiving Steem: exchanging it for another currency or mining it following the bitcoin cryptocurrency model.

The origin of the externally tradeable value of Steem is the first investment from the SteemIt founders. As with BitCoin, the value of each coin theoretically decreases over time as new currency is generated (i.e., as supply increases), but can increase if users find the currency useful or wish to speculate on its future value. It can also be added to by advertising revenue from promoted content [9,10]. Thus, increases in the value of Steem derive from a combination of advertising and currency trading or investments from SteemIt users, currency traders, or others that find other uses for Steem. If new users or advertisers continue to invest in Steem at the same or greater rate than the currency is generated, then there will be no problems. People might buy Steem for speculation purposes if they believe that its value will increase compared to the U.S. Dollar through others investing in it. If this happens then they can convert their Steem back to U.S. Dollars at a profit. People can also invest in Steem to buy Steem Power, which gives them a longer-term stake in SteemIt [11]. If new users and advertisers stop investing in Steem or existing users withdraw their money faster than new users add it then the value of Steem will decrease so that the newest investors will lose most of their money. This is an open Ponzi-like characteristic, and is liable to "pump and dump" investment strategies [12], but the same core model has worked for BitCoin for a long time. SteemIt is protected against internal software-based Ponzi schemes that work on Ethereum [13] by design (only users can own money and not software/wallets). In the final analysis, the ability of SteemIt to reward users for content generation and curation is dependent upon the prevailing financial belief about the likely future exchange value of Steem.

SteemIt's underlying cryptocurrency Steem follows the (so far) successful model of BitCoin through being a blockchain based online currency not backed by governments or commodities [14,15]. The blockchain model secures against hacking by keeping multiple records of transactions and ensuring that the computational power required to hack the currency is prohibitive. Cryptocurrencies are not well understood from an economic perspective because the influence of supply and demand has varied over time for BitCoin [16].

Only one previously published academic paper has reported any analysis of SteemIt: a discussion of software bugs in eight blockchain systems [17]. There is no academic information about site users. This paper investigates SteemIt from the perspective of members' key first step, posting content to the site, to assess what types of content are

valued and which initial posting strategies are the most successful. The aim is to use this as a lever to get insights into how the system works in practice. A secondary goal is to assess the types of post that are considered valuable after the first post. Although it would be possible to analyse all content in the system at once, it is useful to analyse first posts separately. This is because the value of subsequent posts will be partly determined by the size and influence of the poster's following, which itself will be partly due to the success of the first post. Thus, first posts form a more transparent data set to analyse. The main goals of the paper are therefore to discover factors that influence the financial value of a member's first SteemIt post, differentiating them from factors influencing subsequent posts. The goals concern posting rather than other actions in the site (friend networks, offline behaviour, user perceptions and motivations). This is an appropriate starting point because rewarding content creation is at the heart of SteemIt.

Background

Despite the apparent importance of initial steps in a new community, there do not seem to have been any empirical analyses of the first steps taken by users in any online social network. There is knowledge about who joins social networks and the depth of the relationships that are created [18], and recommendations for early actions [19] but little data about initial strategies. Perhaps the closest is an analysis of Facebook that detected behaviours associating with success in ongoing usage of the site. The key feature is the need for ongoing active communication, but falling short of what might be considered excessive communication and friending [20]. Facebook is rarely used to meet new people, probably unlike SteemIt, but is used to communicate with pre-existing friends and to strengthen ties with acquaintances [21]. Thus, Facebook behaviour might not be a suitable guide for SteemIt. The remainder of this review focuses on relevant online social network research.

Online social network dynamics

The creation of friend/follower networks in SteemIt is likely to be important for voting patterns. It is therefore useful to review what is known about online networks in other sites. There are many different types of online network and their dynamics are often emergent from interactions between users rather than transparent from their purpose and affordances. For example, Wikipedia editors implicitly or explicitly collaborate to create high quality content without remuneration. This leads to some counterintuitive effects, such as community stability not being helpful for content quality [22]. A contrasting example is World of Warcraft. Despite being an online game, users commonly formed strong online friendships partly by escaping their offline identities in roleplay [23].

Online social network members often connect to others that are like themselves or have shared interests [24,25]. On some sites, members may spend a substantial fraction of their time browsing the pages of their connections [26], suggesting that connections are important information sources. This explains why information spreads more easily from highly connected users or from people that are otherwise in good network positions [27]. Although social networks can grow in different ways, one successful model is *motivated evangelism*, where existing members get benefits from recruiting new ones [28]. This matches SteemIt on the assumption that a member's followers are more likely to upvote their content.

The typical membership of a social web site is likely to change over time (e.g., [29]). As with any new technology, the first users are likely to be different from others through

being "innovators" or "early adopters" [30] and may also have a different success rate. Higher success may be due to their characteristics or because early adopters have longer to build social capital within a site and may be aided in this by their early joiner status.

Online knowledge sharing and social capital

Many online networks involve an element of knowledge sharing. In the absence of direct financial rewards, users can share knowledge to accumulate collective social capital (i.e., increasing the collective expectation of future help from others: [31]). They may be driven by personal goals, hoping that the system will eventually reward them, or more community-related goals, wanting to improve the community as a public good. They may also support (e.g., upvote) individuals within a knowledge sharing community in the expectation that the individual would be more likely to support them in return [32]. For example, most Reddit upvotes are from users that have not read the post [33], so these may be motivated by prior knowledge of the poster if they are not based solely on the post title.

In addition to social capital, pleasure may also be a motivating factor. In the domain of tourism, people tend to join information sharing online communities to satisfy information needs, but are also motivated by enjoyment and the satisfaction of being part of a community [34].

Yahoo! Answers is another type of online knowledge sharing community and is based on answering questions. This site allows users to post questions in the hope that other members will answer them. The original question posters can then flag appropriate answers [35]. There is no financial incentive to participate but users get public recognition through a virtual reward system. Participants to the similar site AnswerBag.com included topic specialists and people that would gather and cite evidence to produce comprehensive answers [36]. These confirm that financial incentives are not always necessary for content creation and curation, but that recognition may be sufficient.

Financial incentives for information sharing will not necessarily work. Motivation crowding theory argues that they can undermine intrinsic motivation and have a negative overall effect [37]. There is some empirical evidence that financial rewards can reduce intrinsic motivation to share knowledge in some online communities [38] and online in general [39].

Sites like Reddit and Wikipedia confirm that people can freely share knowledge without payment, but it remains to be seen whether SteemIt's introduction of micropayments as a core incentive for knowledge sharing undermine this paradigm or whether users ignore this incentive because other factors motivate them.

Social news sites: Reddit and Digg

The social news website Reddit is a highly successful non-commercial model for SteemIt (9th most visited website according to Alexa.com in July 2017: http://www.alexa.com/siteinfo/reddit.com; see also [40]). Reddit users can contribute content, called "self-submissions", or URLs. If other users see and like the post then they can upvote it, with the most popular being displayed on the Reddit homepage. This acts like a filter to get the most interesting news at the top [41]. People that provide the content or links are rewarded by karma points or gold "creddits", both with no financial exchange value, whereas people that look through new content to upvote interesting posts (curation on SteemIt) are not rewarded. Although Reddit is a highly popular site of its type, relevant content is often ignored. For example, half of the site's top content was initially overlooked

[42]. Like SteemIt, Reddit organises posts by topic (subreddits). Reddit is financed by banner advertising and promoted links. There are occasional tensions between users and the owners of the site, leading in one example to a strike, the July 2015 blackout, against an unpopular policy [46]. Other sites, such as Imgur, have a broadly similar organisation for different types of content.

Reddit has grown dramatically in the past few years after a long period as a smaller community since its formation in 2005. Using Google Trends search data as a proxy for its popularity, its growth was slow from 2005 to 2010 but it has since grown linearly (https://trends.google.co.uk/trends/explore?date=all&q=reddit). During this period, the character of Reddit has changed with the introduction of a huge number of subreddits (125,662 by December 2012), an increase in the proportion of self-submissions from about 1% of all posts in 2008 to over a third in 2012, and an increase in the proportion of images submitted, with multimedia (mainly images and videos) forming about half of all submissions at the end of 2012 [44]. SteemIt seems to be following the trajectory of Reddit by focusing on content rather than posting links to external sites.

Reddit has become a significant repository of citizen journalism and has provided a substantial alternative source of information to the mainstream media for some events, such as the Boston Marathon bombing, but its coverage seems to be similar overall for important events [45]. It is also a source of political deliberation [46,47] and educational resources [48]. Some Reddit users spend substantial amounts of time interacting on the site, to the extent that they only give up when they are tired [49].

Digg.com was founded in 2004 as a social news website. It has since been moderately successful (Alexa rank 1990 in July 2017, http://www.alexa.com/siteinfo/digg.com) although the digg.com website is now a news portal with curated main pages. A study of the former social news version of Digg found that the degree of interactions between users on the site was much greater in narrow topics than in more diverse categories [50], suggesting that member interactions are important but not essential.

Other reward models

The Stack Exchange suite of websites have a similar reward mechanism to Reddit, but with the goal of answering questions. Users can post and answer questions, being rewarded with reputation by other users for good answers [51,52]. Visible Stack Exchange rewards, such as reputation badges (e.g., 'Custodian', 'Explainer', 'Curious'), are an effective non-financial motivation that increases average user activity [53]. The system is very successful although users tend not to be very careful when selecting good answers to vote for [54]. Although based on question answering, Stack Exchange forums generate a body of useful knowledge, typically focused around a topic. Its badges are a form of gamification because they challenge (receptive) users to carry out the tasks required to achieve them [53].

An alternative reward system is Flattr (a Scopus search for this term returned no hits). It allows users to reward sites that they like by adding them to a list to share the individual's monthly Flattr payment. This content is on sites other than Flattr, however, differing from SteemIt.

In summary, for all the models discussed in this section, individual users vote to reward good content creation. For most (except SteemIt, Flattr) the reward is intangible in the form of visible markers of reputation or success (high post upvotes, badges achieved) and the outcome is a widely useful resource. The purpose differs between news plus

general content (SteemIt, Digg.com, Reddit), question answering (Stack Exchange), and general (Flattr) sites. Of these, SteemIt is the only site backed by its own cryptocurrency. Of course, other cryptocurrencies may be used to reward good content on an ad-hoc basis (e.g., Dogecoin) but there do not seem to be other successful attempts to make this process systematic.

Factors affecting the success of a post

The likelihood of a post being successful can be predicted to some extent from its properties or those of its poster in some online environments. Background information on this is useful to give context to analyses of SteemIt. The most relevant information in most contexts is about the popularity and network of the poster (e.g., [55]) but this is not relevant to first posts. Some topics are orders of magnitude more likely to be retweeted than others [56], making this perhaps the most significant single content attribute for post popularity.

The emotional state of a Twitter user affects the likelihood that they retweet [57]. The presence of positive or negative sentiment all associate with a slightly increased number of favourites [58]. The sentiment content of a tweet is also thought to be useful to predict retweets [59,60].

Longer answers in Stack Exchange are more likely to be selected as the best answer to a question [54] and longer tweets are slightly more likely to be favourited [58]. The relationship between tweet length and retweeting is more complex, however [61].

For news tweeted by major organisations, time of day can affect the number of retweets as well as the number of people receiving it at an early stage [62]. Time since posting impacts upvotes in Reddit and similar sites [63]. The presence of hashtags or URLs also associates with increased retweeting [56,61]. On Reddit, early upvotes generate a larger final number of upvotes [64] and so there is a degree of imitation amongst voters. The design of a system also impacts the posts that a user sees [65] so the findings in this section do not necessarily transfer to SteamIt.

Research questions

Guided partly by the above background and partly by the research goals, the following research questions drive this study.

- RQ1: Do post length, date, sentiment and topic influence the financial value of a member's first SteemIt post?
  - o Post length can influence popularity, with longer posts presumably tending to carry more information. Nevertheless, shorter posts may embed images or videos.
  - o First post date is important because early users may be a different type of person (e.g., an "innovator" or "early adopter": Rogers, 2010) than later members. This seems particularly relevant to a reward-based site.
  - o Sentiment is known to affect the success of users and posts in the social web.
  - o Post topic influences social web popularity and is an obvious angle to investigate given that SteemIt users can post about anything.
- RQ2: Do different factors influence the financial value of members' subsequent SteemIt posts?

Social capital features are noticeably missing from the research questions because of the focus on first posts, which will be predominantly from users starting out in the network.

Time of day is also not included since SteemIt is an international platform and so there is no baseline time to work from.

Methods

The research design was to extract all SteemIt comments, separate first and subsequent posts in the most common language, and investigate the linguistic and length properties of these posts quantitatively, using statistical methods where possible. A quantitative approach was chosen rather than a more qualitative method, such as a content analysis of a random sample of posts, due to the large volume of data and the difficulty in getting a useful random sample in the context of multiple potential topics and highly skewed reward amounts per post. Post date, length, sentiment and topic were selected because these are amenable to a quantitative analysis. Other aspects may be important but were ignored for the primary methods.

The complete contribution database of SteemIt was downloaded on July 24, 2017 directly from its public database using remote SQL access [66]. Each record includes the initial creation date, the number of SteemIt Dollars earned, the author's SteemIt name, and the language of the post, as estimated by the SteemIt language detection algorithm. This step produced 4,795,805 records, comprising blog posts, comments and replies.

A set of users' first posts was created for an analysis of introductions. For this, the first contribution of each user was extracted, a total of 96,792. The 20% of first posts not in English were then removed to ensure a more homogeneous dataset for the analysis and particularly for the word frequency topic analysis, leaving 77,905. The 19,558 first contributions that were comments on posts or replies rather than posts were also removed, giving 58,347 first contributions that are English blog posts. After removing a few blank contributions, the final count was 58,054. This process ignores people with a first contribution that was a comment or reply rather than a blog post to ensure that no posts have a hidden advantage from the author first building up a following from commenting on others' posts.

A set of subsequent English posts was then extracted to analyse longer term contributions to the site. For this, all non-English posts were removed from the initial set of 4,795,805 records, leaving 3,925,664 English blog posts and comments (82%). All 2,915,313 comments and replies were removed, leaving 1,010,351 English blog posts. From this reduced set, the first English blog post of each author was removed (74,718), keeping the remaining 935,633 English blog posts by authors that had previously authored an English blog post. After deleting empty posts, this left 925,092. For authors with a first post not in English, this would also eliminate their first post in English. This is a conservative step to reduce the risk that subsequent posts had an introductory role for a different language community, such as through being an English translation.

The comments were segmented by month for most analyses to reduce the risk that changes over time would be conflated with other changes.

The average financial value of each post was plotted over time in conjunction with the number of posts to detect a relationship between the two. The amount of Steem Dollars per post was highly skewed (within-month skewness between 6 and 61 after the first month) and so geometric means were used instead of arithmetic means to report the average financial value of comments. Since there were many posts with 0 Steem Dollars, following standard practice, 1 was added to each value before calculating the geometric mean and then subtracted afterwards. Although votes could be counted instead of Steem

Dollars, some votes are apparently made by automated processes on a large scale and so vote counts can be misleading. The process that translates votes into Steem Dollars reduces the impact of spam-like upvoting by reducing the value of multiple upvotes from the same user.

Post length was correlated with Steem Dollars to detect an overall relationship. Spearman correlations were used instead of Pearson correlations because of the skewed nature of the data.

The average positive and negative sentiment strengths of posts were compared against the amount of Steem Dollars received to check whether there was a relationship between the two. Sentiment was detected using the software SentiStrength that estimates the strength of positive and negative sentiment in a text using a dictionary of sentiment terms and additional linguistic rules [67]. Its outputs are on a scale of 1 (no positive sentiment) to 5 (very strong positive sentiment) and -1 (no negative sentiment) to -5 (very strong negative sentiment). Although there are many other sentiment analysis programs (e.g., [68]), SentiStrength is appropriate for having proven high accuracy over many different types of social web text [69]. For the comparison, Steem Dollars were bucketed (e.g., 1 cent, 2 cents, 3-4 cents, 5-8 cents, 9-16 cents) and the arithmetic mean number of dollars per post calculated for each one, together with a bootstrapped 95% confidence interval. Although the dataset is complete rather than a random sample and so confidence intervals are inappropriate, they are useful to delimit the likely underlying trend. This assumes that each post is a randomly selected post from the theoretically infinite space of posts that could plausibly have been written (i.e., an apparent population [70]).

High and low value topics were detected using a word frequency approach rather than topic modelling [71, e.g., 72] because word frequency data is more transparent and the key issues are not necessarily topics in the traditional sense, as evident from the results. Words associating with higher value posts were detected by splitting the dataset into two: initial posts that attracted 0-9 cents and posts that attracted 10+ cents. This cut off was chosen heuristically. Although 10 cents may be a very small amount of money to most people using the system, only 12.8% of first posts earned more and so it is large enough to give a useful split between more and less successful posts, without being so low as to be irrelevant. Chi-square tests were then conducted for each word in each post to test whether it occurred disproportionately often in the 0-9 cent group or in the 10+ cents group. The chi-square test assesses the statistical significance of the difference between the two percentages. So, for example, if the word *friend* occurs in 26.5% of the posts valued at 10+ cents but only 15.0% of the posts valued at 0-9 cents then the chi-squared statistic tests how likely this outcome is if the underlying likelihood (using the apparent population idea again: [70]) that *friend* occurs in a post is independent of whether the post is worth less than 10 cents.

The 20 words with the highest chi-square value for each set were analysed in more detail. This deeper analysis was conducted by reading a sample of posts containing the term and finding words associating with the combination of the term and the financial group, again using the chi-square test. For example, the top word for the 10+ group was *SteemIt* (chi-square 1993.2). The word that most strongly associated with *SteemIt* in the 10+ group was *community* (chi-square 2037.8 for SteemIt, 10+ cents), with 40.7% of posts in the 10+ group containing *SteemIt* also containing *community*.

The average value of a term (or sentiment range) was estimated by calculating the geometric mean value of all posts containing that term. Geometric means were again used due to the high level of skewing in the data.

First post results

The results are reported separately for each key aspect analysed. Overall, the 58,054 first posts attracted a geometric mean of 1.22 cents each, with a 95% confidence interval of (1.19, 1.26). The highest individual post value was $20,680.83 for an article about promoting SteemIt in South Africa[2]. Over four fifths of first posts (82.2%) had a value of 0 cents.

Post date

For most of the time the average reward earned by a first English post has been a few cents (Figure 1). The main exception is that the early posts from up to June 2016 attracted a substantially larger average reward. A secondary exception is February-March 2017. In both cases these exceptions were followed by rapid expansion in the system. Unless this is a coincidence, there are at least two plausible explanations. First, system expansion might generate a pool of new users to vote for existing content more rapidly than it generated new content to be voted for. Second, high quality content or active recruiters might attract new users who vote for the existing users or content. The former case suggests an advantage from joining the site just before it expands rapidly.
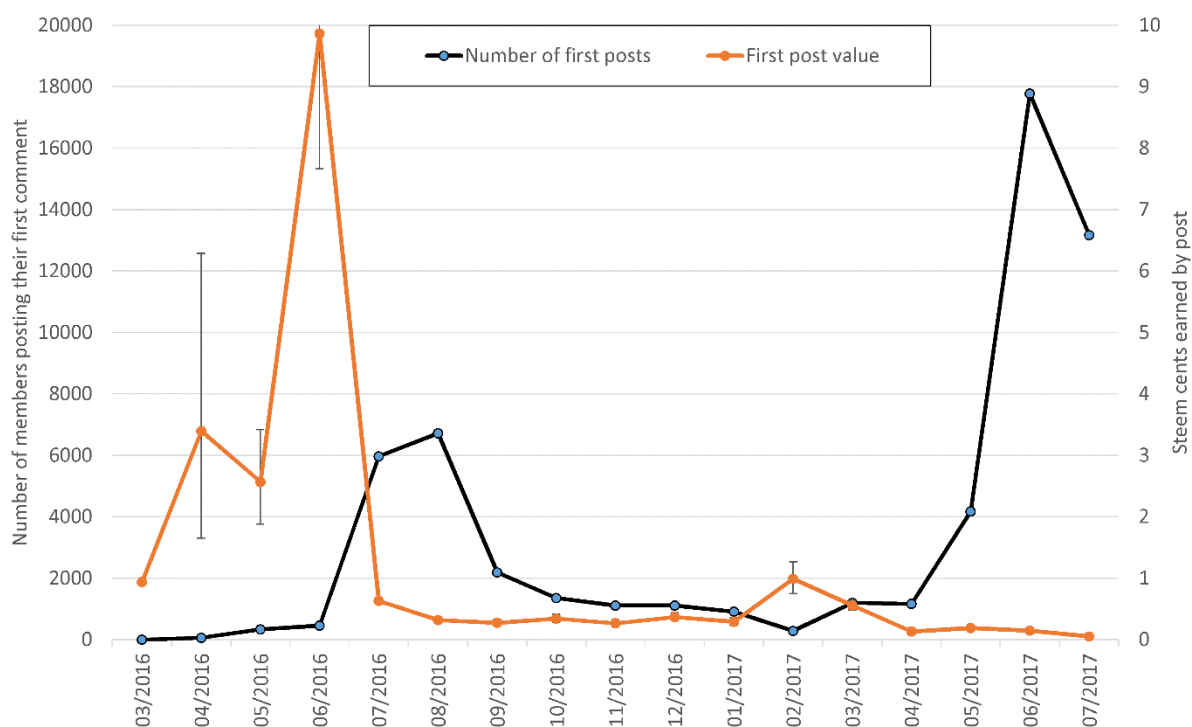

Figure 1. The average (geometric mean) number of Steem Dollars received by users' first English comments against the month in which the comment was first posted. Error bars show 95% confidence intervals.

---

[2] https://steemit.com/money/@steemdrive/sa-s-city-of-durban-to-be-engulfed-by-steem-the-world-to-follow

Length

Longer posts tend to attract more Steem Dollars, although the relationship has decreased dramatically over time. It has been very low since August 2016 and is now effectively non-existent (Figure 2). It is possible that competition from the large number of users has reduced the effectiveness of longer posts and that members now prefer shorter posts that can be read and assessed/curated more quickly. Longer posts at earlier dates presumably had more content and therefore either displayed more seriousness or more valuable information.
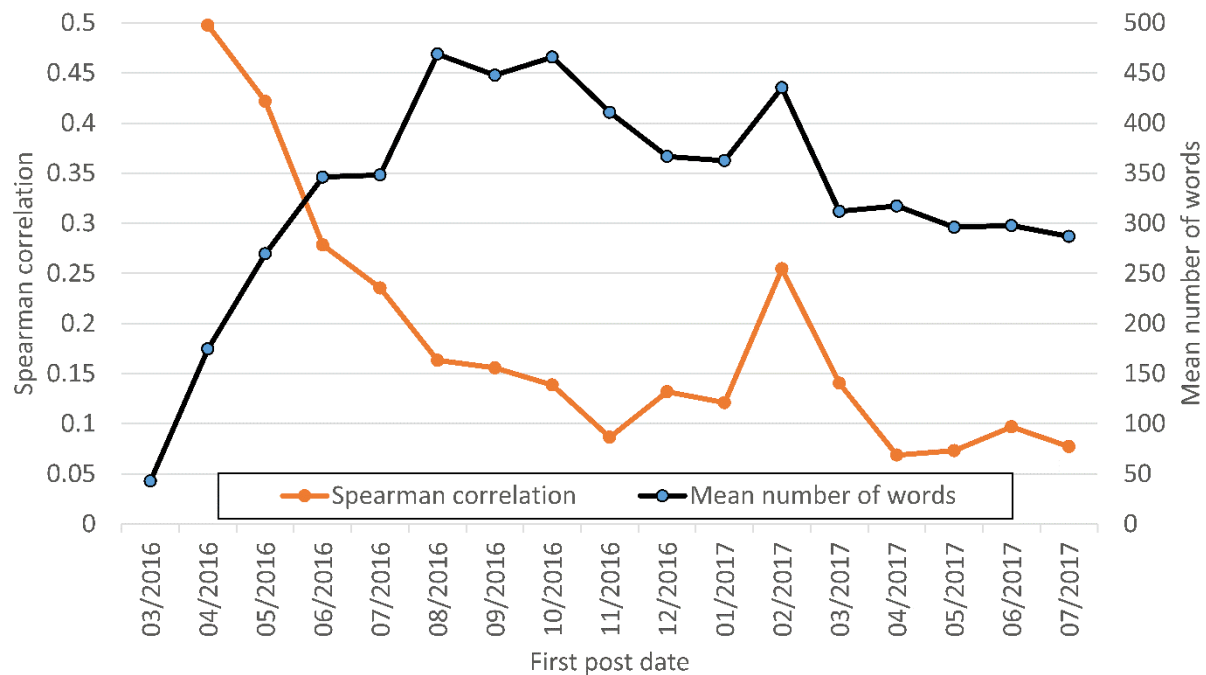

Figure 2. Spearman correlation between the length of a post (number of words) and the number of Steem Dollars received by publication month for users' first comments.

Sentiment

First blog posts in SteemIt were moderately positive (an average positive sentiment strength 2.81, 95% CI [2.81, 2.82]) and mildly negative (average negative sentiment strength 2.42, 95% CI [2.41, 2.43]) (Figure 3). The overall average (geometric mean) value per post is 1.22 Steem cents. For posts with moderate or strong positive sentiment (value: 3, 4 or 5) and no or weak negative sentiment (value -1 or -2) the average is 1.47c, 95% CI (1.40, 1.55). In contrast, the average for posts with moderate or strong negative sentiment (-3, -4, or -5) and no or weak positive sentiment (1 or 2) the average is only 0.56c (0.49, 0.63), and for posts with no sentiment of any kind the average is even lower, at 0.32c (0.28, 0.36). Hence, exclusively positive sentiment is an advantage and exclusively negative sentiment or no sentiment is a disadvantage. Nevertheless, the combination of moderate or strong positive (3, 4 or 5) and negative sentiment (-3, -4, or -5) is an even bigger advantage, with an average value of 1.77c (1.68, 1.85). Such posts may be balanced arguments for a controversial topic, for example, containing both positive and negative points. Balanced arguments may be more respected and posts that are positive and enthusiastic may also be appreciated.
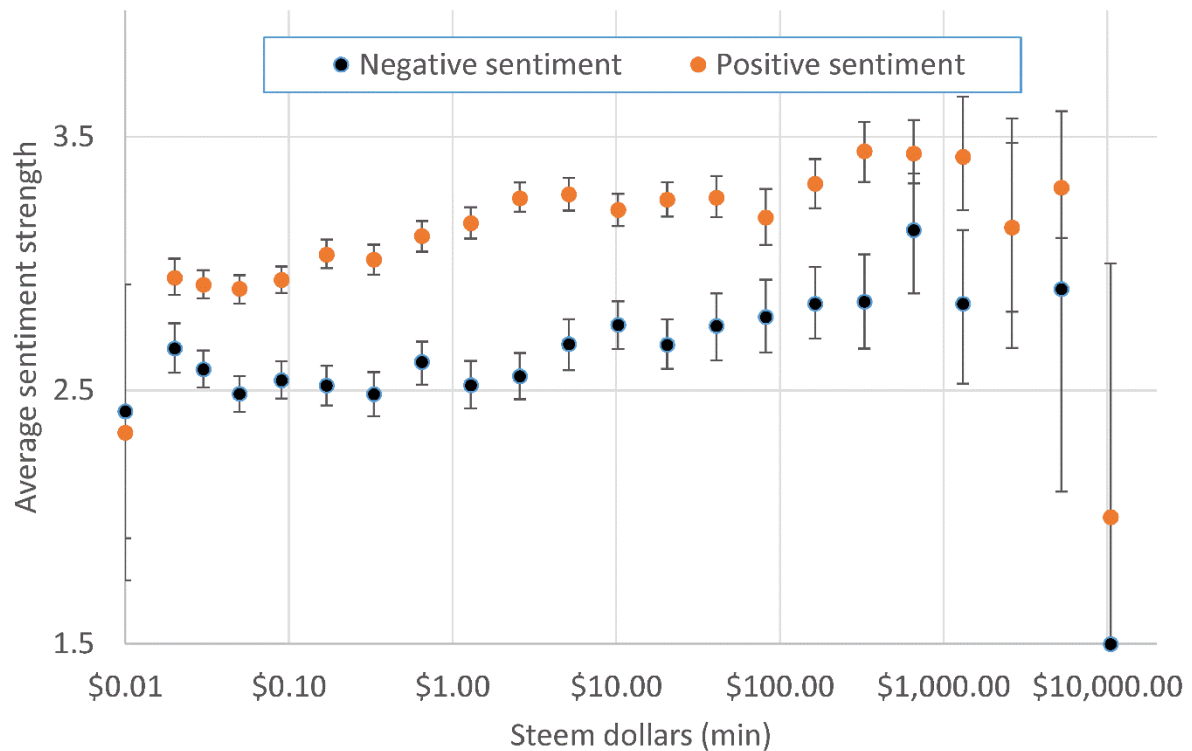
Figure 3. Average negative and positive sentiment strengths of first English posts against the number of Steem Dollars earned. The data is bucketed into buckets of doubling size.

Topic

The words associating with higher values (at least 10 cents) mainly relate to giving personal information, as indicated by pronouns and *name*. The terms *community* and *SteemIt* were often used in texts expressing pleasure at joining the SteemIt community or using phrases like, "hello SteemIt community". The term *year* is typically used in self-descriptions, giving the year in which important events happened. This suggests that introductions are valued within the site.

Although not evident in Table 1, many posts requested friendship in the site and so this was also examined. Perhaps surprisingly (e.g., tweets containing #followyouback tend not to be retweeted: [60]), posts that contain the request, "follow me" earn substantially above average, at 3.04c (2.62, 3.51).

Table 1. The 20 words that tend to occur most in first posts in English accruing at least 10 cents compared to their frequency in posts earning 0-9 cents, as judged by the chi square statistic. For example, 43.7% of posts with at least 10 cents contained *SteemIt*, compared to only 20.3% of posts with 0-9 cents.

Word	10+ cents	0-9 cents	Posts	Chi square	Average post value (geo. mean cents)
SteemIt	43.7%	20.3%	13505	1993.2	3.60 (3.39, 3.82)
my	74.1%	50.2%	30914	1486	2.06 (1.97, 2.14)
me	58.7%	36.5%	22848	1343.7	2.33 (2.22, 2.44)
name	35.0%	17.9%	11682	1168.3	3.06 (2.86, 3.26)
I	82.4%	62.9%	37944	1088.6	1.75 (1.68, 1.81)
community	26.0%	12.2%	8134	1025.3	3.64 (3.36, 3.93)
am	45.2%	27.2%	17139	1010.1	2.44 (2.31, 2.57)
myself	25.3%	12.3%	8108	917.1	3.38 (3.12, 3.64)
year	44.5%	27.6%	17272	888.6	2.31 (2.24, 2.50)
about	58.6%	40.4%	24803	875.5	2.05 (1.95, 2.14)
love	32.4%	17.9%	11451	871.8	2.77 (2.59, 2.96
I'm	33.4%	18.9%	12052	832	2.64 (2.48, 2.81)
post	33.6%	19.2%	12211	814.7	2.71 (2.55, 2.89)
here	46.5%	29.9%	18608	812.6	2.26 (2.15, 2.38)
in	85.2%	71.1%	42303	658.5	1.56 (1.51, 1.62)
and	92.3%	80.1%	47392	653.2	2.04 (1.93, 2.15)
friend	26.5%	15.0%	9563	623.7	2.75 (2.56, 2.96)
like	53.2%	38.2%	23274	611	1.94 (1.85, 2.03)
so	58.0%	42.8%	25966	608.5	1.88 (1.80, 1.96)
of	86.6%	73.6%	43689	593.1	1.53 (1.48, 1.58)

In contrast to the high frequency words associating with increased value (Table 1), the words that associate with low value are much rarer and with varied reasons (Table 2). Common themes in the table include giving diet advice or a recipe, discussing cryptocurrency issues, such as trading or competitors, or religion, sports and news. The generic term *sport* [2.31 cents (1.94, 2.72)] has an above average value, but mostly due to its inclusion in descriptions of the types of sport that the member enjoys playing or watching. Similarly, the general term *religion* [1.35c (1.04, 1.70)] has above average value, mostly from use in self-descriptions or abstract discussions covering issues such as problems caused by religion, tolerance of religious beliefs or differences between religions.

Table 2. The 20 words that tend to occur most in first posts in English accruing under 10 cents compared to their frequency in posts earning 10+ cents, as judged by the chi square statistic.

Word	0-9 cents	10+ cents	Posts	Chi square	Comment and average cent value (95% CI)
percent	1.4%	0.8%	750	14	Financial and SteemIt news 0.10 (0.10, 0.10)
carbohydrate	0.2%	0.0%	113	12.4	Diet advice 0.08 (0.01, 0.16)
segment	0.4%	0.1%	221	12.2	Financial or consumer news 0.12 (0.06, 0.19)
president	1.8%	1.2%	991	11.2	Discussion of U.S. elections 0.10 (0.10, 0.10)
Christ	0.5%	0.2%	271	10.4	Religious introduction or sermon 0.12 (0.06, 0.18)
accs	0.1%	0.0%	69	10.2	Selling fake SteemIt bot votes 0.12 (0.06 0.18)
suspected	0.2%	0.1%	128	9.1	News about crime or illness 0.10 (0.02, 0.18)
according	4.3%	3.6%	2461	8.9	Various news reports 0.10 (0.09, 0.12)
trump	1.4%	1.0%	775	8.7	Discussion of U.S. elections 0.10 (0.09, 0.12)
reduced	0.8%	0.5%	437	8.2	Various news reports 0.10 (0.09, 0.12)
victory	0.6%	0.3%	318	7.9	U.S. election result 0.08 (0.03, 0.12)
NFL	0.2%	0.1%	119	7.9	NFL preseason news 0.02 (0.00, 0.05)
calories	0.4%	0.2%	227	7.8	Diet advice 0.07 (0.02, 0.11)
fatty	0.2%	0.0%	92	7.5	Diet advice about fatty acids 0.09 (0.01, 0.18)
Ronaldo	0.1%	0.0%	51	7.5	News about footballer 0.08 (0.00, 0.20)
Nigerian	0.3%	0.1%	149	7.4	Nigerian news 0.08 (0.00, 0.20)
mankind	0.6%	0.3%	302	7.3	Environment or religion discussions 0.08 (0.00, 0.20)
price	7.0%	6.1%	3989	7.3	Currency trading and financial news 0.10 (0.09, 0.11)
depart	0.1%	0.0%	64	7.3	Bible and stories 0.11 (0.00, 0.25)
ICO	1.4%	1.0%	806	7.2	New cryptocurrency sale 0.11 (0.00, 0.25)
wireless	0.2%	0.1%	124	7.1	Communication technology news 0.05 (0.00, 0.11)

Subsequent post results

Excluding first English posts, the 925,092 non-empty subsequent posts in English earned an average of 2.01 cents each, with a 95% confidence interval of (1.99, 2.02). The highest individual post value was $44,529.19, an article from July 27, 2016 about the first graphical user interface for SteemIt[3]. The higher value of subsequent posts compared to first post may be due to successful users being more active or subsequent posts attracting votes from followers of the author accumulated from earlier posts. Just under four fifths of subsequent English posts (78.2%) earned nothing.

Post date

The average reward earned by subsequent English posts is slightly more than for first posts overall, except in the early months (Figure 4). The overall reward pattern is broadly like that for first posts (Figure 1).

---

[3] https://steemit.com/piston/@xeroc/piston-web-first-open-source-steem-gui---searching-for-alpha-testers
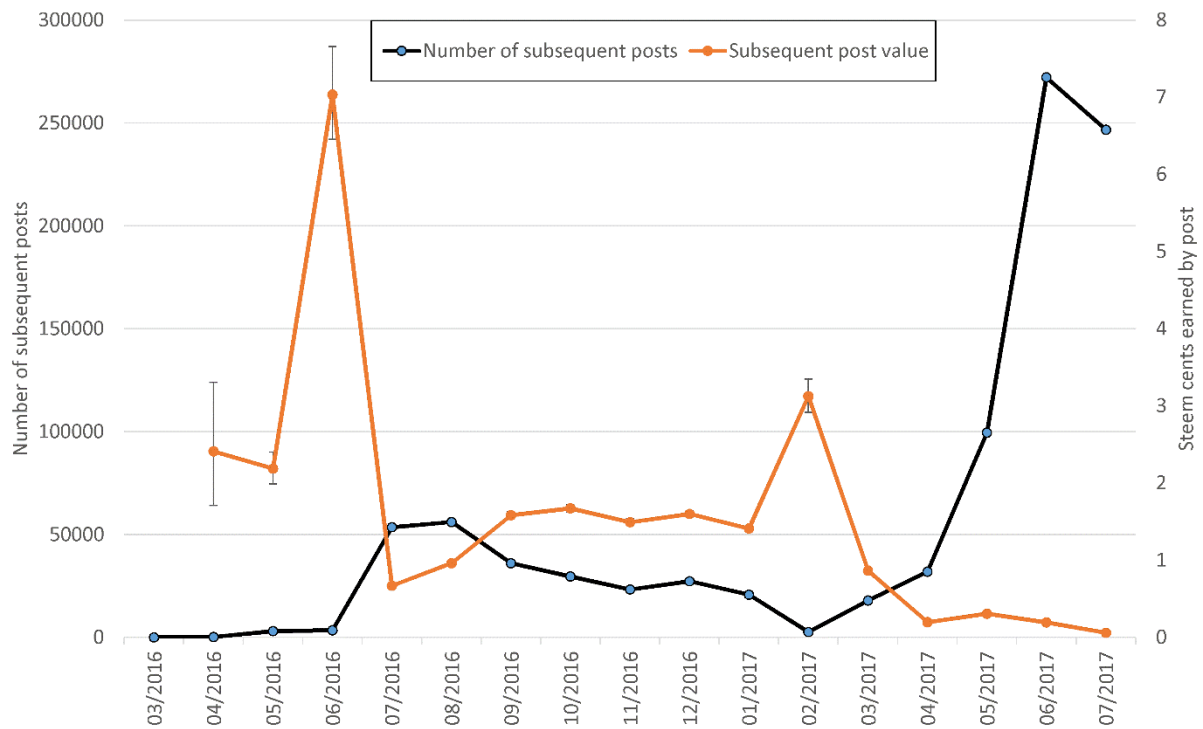
Figure 4. The average (geometric mean) number of Steem Dollars received by users' English comments after their initial post against the month in which the comment was first posted. Error bars show 95% confidence intervals.

Length

Longer posts tend to attract more Steem Dollars, although the relationship may be decreasing over time (Figure 5). The correlations are low to moderate, so post length has a small but non-negligible influence on the number of Steem Dollars earned. This influence is slightly stronger overall than for first posts.
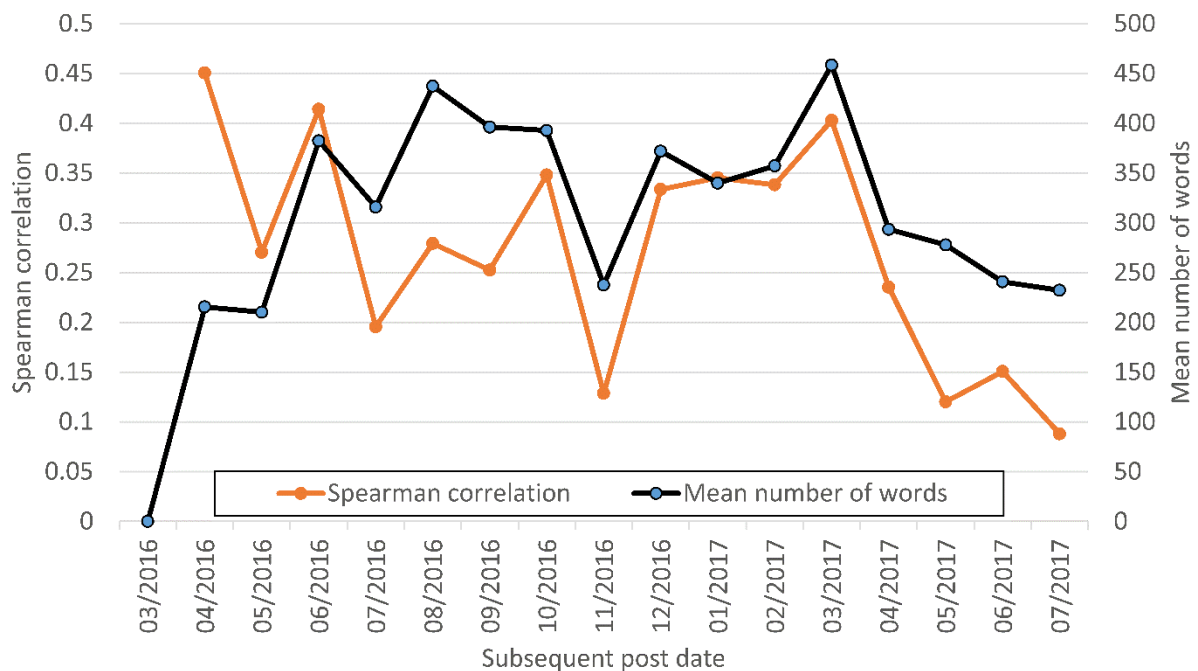
Figure 5. Spearman correlation between the length of a subsequent English post (number of words) and the number of Steem Dollars received by publication month for users' first comments.

Sentiment

Subsequent blog posts in SteemIt were mildly positive (an average positive sentiment strength 2.37, 95% CI [2.37, 2.37]) and mildly negative (average negative sentiment strength 1.61, 95% CI [1.60, 1.61]). Both positive and negative sentiment were weaker than for first posts, especially for posts earning little money (Figure 6). For posts with moderate or strong positive sentiment (value: 3, 4 or 5) and no or weak negative sentiment (value -1 or -2) the average value 0.35c, 95% CI (0.35, 0.36) is *below* that for all comments (0.42c). In contrast, the average for posts with moderate or strong negative sentiment (-3, -4, or -5) and no or weak positive sentiment (1 or 2) the average is higher 0.41c (0.41, 0.42) and about the same as the all-posts average, and for posts with no sentiment of any kind the average is low at 0.17c (0.17, 0.17). Hence, exclusively positive sentiment is a disadvantage for subsequent posts and exclusively negative sentiment or no sentiment is a disadvantage. Nevertheless, the combination of moderate or strong positive (3, 4 or 5) and negative sentiment (-3, -4, or -5) is a big advantage, with an average value of 2.40c (2.38, 2.43).

Thus, the importance of combining positive and negative sentiment is the same for first and subsequent posts, only first posts benefit from focusing on positive sentiment.
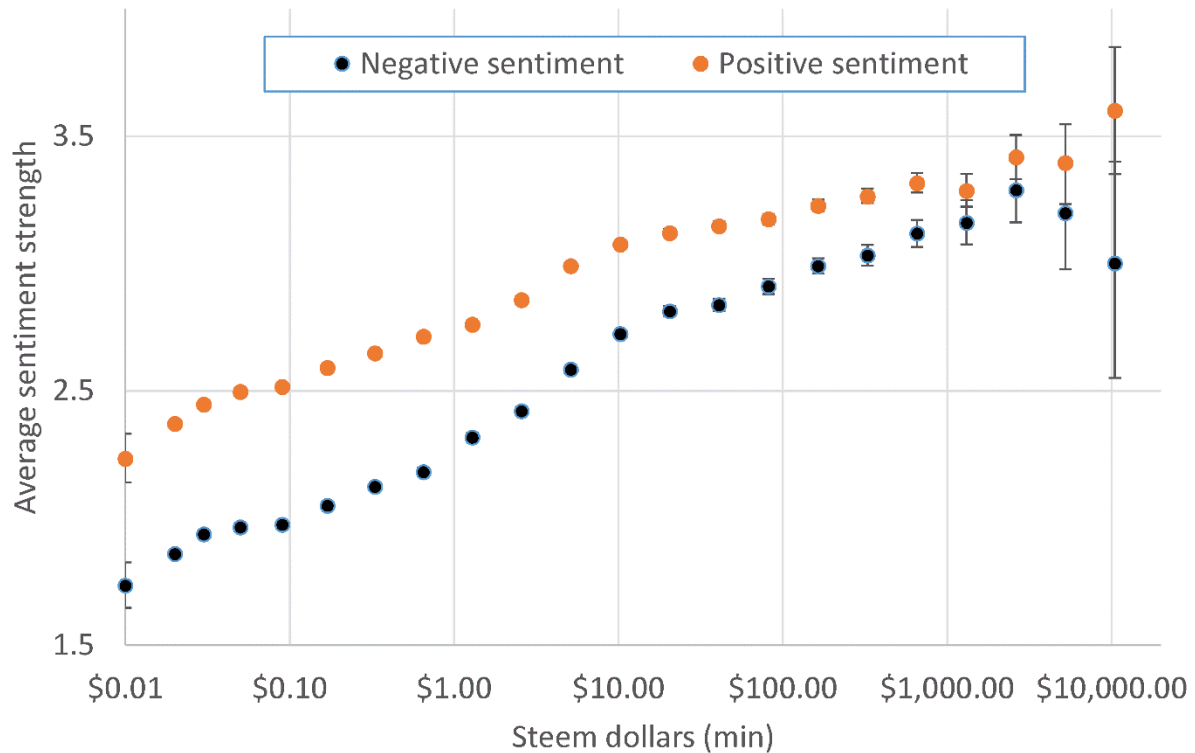
Figure 6. Average negative and positive sentiment strengths of subsequent English posts against the number of Steem Dollars earned. The data is bucketed into buckets of doubling size.

Topic

The words associating with higher value (at least 10 cents) subsequent posts tend to be function words (Table 3). This suggests that post length is more statistically important than post topic for subsequent English posts. The exceptions are SteemIt and personal pronouns (I, my, me, you), partly reflecting the trend for first English posts (Table 1).

Table 3. The 20 words that tend to occur most in subsequent English posts accruing at least 10 cents compared to their frequency in posts earning 0-9 cents, as judged by the chi square statistic.

Word	0-9 cents	10+ cents	Posts	Chi square	Average post value (geo. mean cents)
I	72.0%	50.2%	498365	25016.3	1.75 (1.68, 1.81)
my	58.3%	36.9%	374836	24882	2.06 (1.97, 2.14)
SteemIt	26.5%	11.6%	130540	24169.8	3.60 (3.39, 3.82)
post	33.4%	16.7%	180488	23276.9	2.71 (2.55, 2.89)
so	52.1%	33.0%	335128	20689.5	1.88 (1.80, 1.96)
me	46.6%	28.3%	290973	20298.1	2.33 (2.22, 2.44)
on	70.3%	51.0%	502167	19617.2	1.64 (1.58, 1.71)
you	71.0%	51.9%	510348	19257.6	1.65 (1.59, 1.72)
some	44.1%	26.6%	273369	19234.6	1.91 (1.82, 2.01)
if	51.9%	33.4%	338133	19214.1	1.65 (1.57, 1.73)
like	49.1%	31.0%	315549	19065.8	1.94 (1.85, 2.03)
with	68.5%	49.5%	488139	18788.2	1.69 (1.62, 1.74)
but	55.0%	36.5%	366805	18609.9	1.80 (1.72, 1.88)
for	76.2%	57.8%	563739	18515	1.62 (1.56, 1.69)
about	46.2%	28.8%	294056	18297.2	2.05 (1.95, 2.14)
out	42.6%	25.7%	264189	18231.9	1.81 (1.72, 1.90)
have	59.9%	41.4%	411803	18136.9	1.75 (1.68, 1.82)
this	73.6%	55.3%	540426	17918.1	1.62 (1.56, 1.68)
all	53.0%	35.0%	352019	17908.9	1.79 (1.71, 1.87)
be	60.2%	41.9%	416275	17736.3	1.70 (1.63, 1.77)

The terms associating with low value subsequent English posts tend to be automated or semi-automated repeated posts for financial or currency trading news, competitions or scams (Table 4).

Table 4. The 20 words that tend to occur most in subsequent English posts accruing under 10 cents compared to their frequency in posts earning 10+ cents, as judged by the chi square statistic.

Word	0-9 cents	10+ cents	Posts	Chi square	Comment and average cent value (95% CI)
SegWit2x	0.2%	0.0%	1351	203.6	Bitcoin change proposal 0.15 (0.10, 0.21)
@bguerrero1986	0.1%	0.0%	1060	196.5	Automated Bitcoin news
Aceh	1.2%	0.8%	10509	161.3	Indonesian news and info. 1.00 (0.93, 1.07)
colorchallenge	0.2%	0.0%	1450	139.3	Competition entries 0.40 (0.30, 0.50)
June	2.2%	1.8%	20041	138.3	Cryptocurrency news in June 1.61 (1.53, 1.70)
fasl	0.1%	0.0%	651	126.5	Fully Automated Steem Lottery 0.03 (0.00, 0.05)
@kalemandra	0.2%	0.1%	2067	126.3	Colourchallenge competition entries 0.58 (0.47, 0.69)
adress [sic]	0.1%	0.0%	1111	111.3	Motorbike video channel repeated posts 0.40 (0.29, 0.52)
@murh	0.1%	0.0%	674	106.6	Fully Automated Steem Lottery 0.14 (0.08, 0.21)
xrp	0.2%	0.1%	1772	103.2	Cryptocurrency news 0.64 (0.52, 0.77)
segwit	0.4%	0.2%	3508	100.8	Bitcoin change proposal 0.88 (0.77, 1.00)
amid	0.3%	0.1%	2290	97.9	Financial news 0.78 (0.65, 0.91)
Indonesia	0.8%	0.5%	6771	95.7	Indonesian news and info. 1.17 (1.07, 1.28)
@fasl	0.1%	0.0%	481	87	Fully Automated Steem Lottery 0.08 (0.01, 0.15)
@gauravchugh	0.1%	0.0%	428	83.1	Short image posts 0.04 (0.02, 0.11)
BitConnect	0.2%	0.1%	1260	80.3	Currency exchange news 0.53 (0.41, 0.66)
ini	0.2%	0.1%	1795	79.1	Indonesian/English bilingual posts 0.75 (0.61, 0.90)
JVZoo	0.1%	0.0%	408	76.6	Scam messages 0.04 (0.00, 0.09)
#colorchallenge	0.1%	0.0%	964	76.6	Competition entries 0.40 (0.28, 0.52)
u	0.8%	0.6%	7001	74	Informal language 1.35 (1.23, 1.47)
BIP	0.1%	0.0%	1074	73.9	Bitcoin Improvement Proposals 0.60 (0.45, 0.78)

Discussion and limitations

The quantitative approach used here is a limitation because it does not consider the influences or context of each post, such as whether the poster was bringing a pre-existing audience with them, whether the content of the post was novel, and whether it was well written. The analysis implicitly assumes that all posts were genuine attempts to create useful content and so it is reasonable to assess their value, although some posts may have been tests and others may have been automatically generated.

The relatively low average rewards of first (1.22 cents) and subsequent (2.01 cents) posts in English suggests that this **financial incentive is too low to be a motivating factor** for typical users, unless they can produce non-trivial posts rapidly enough for their goals. Nevertheless, some posts reap substantial rewards and so there may still be a lottery-like incentive for people that hope to get a jackpot with some posts. Moreover, since many people use sites like Reddit without pay, the low reward is not necessarily a disincentive but may instead be an extra to an activity that the user enjoys or is prepared to engage in without financial reward. Since the likelihood of getting upvotes for a post probably increases with the number of followers in SteemIt, users may also enjoy the challenge (like game playing: [73]) of increasing their follower count with the prospect of increasing their future rewards.

The relative success of personal introductions for SteemIt first posts suggests that **SteemIt members widely recognised the importance of social capital** [74]. This makes sense because reciprocal behaviour is mutually and immediately financially beneficial on SteemIt, albeit for very small amounts of money. Thus, at least when joining SteemIt, social capital (in the form of networking) rather than content creation is the driving factor. This conclusion is supported by the importance of the affective dimension of posts, with positive sentiment associating most strongly with increased financial reward. This fits with previous research about the importance of interactions in Digg [50] and perhaps Reddit (for politics: [46]).

The continued importance of the personal dimension and the SteemIt platform for subsequent English posts (Table 3 and the sentiment results) suggests that **social capital is more important than content for subsequent posts**. This is more worrying given that the rationale for the payment for upvotes is that it is to reward users for creating content. Thus, it seems possible that users overall are more focused on social capital than content, even after their initial introductory posts. Nevertheless, browsing the site shows that many users create interesting and innovative unique content, such as themed sets of pictures, that attracts upvotes worth a few dollars. It is not clear from the quantitative approach here whether users typically combine quality content with social capital elements, such as explicit requests for upvotes, or personal information about the content (e.g., that it is a hobby) or the provenance of the content (e.g., "a photo I took today").

Despite the importance of SteemIt for higher value posts, topics related to finance and cryptocurrency were found in lower value posts (Tables 2, 4). This suggests that **members are not strongly interested by immediate financial gain** or that there is too much of this content on SteemIt for the members, or that a high proportion of the content is low quality. It is possible that the few members that have made a substantial amount of money from the system may be more interested in these post topics.

Conclusions

The focus on personal aspects and the SteemIt site in the more successful posts suggest that the SteemIt model has led to a focus on social capital in the form of network building for new and, to a lesser extent, continuing users. This seems to be more important than content creation, which may be a long-term problem for the site unless it can generate enough useful content to sustain its original pay-for-content model. This issue is exacerbated by the very low typical value of individual posts. Content generation may still occur either because users do not care about financial incentives (as for Reddit) or because they are prepared to make a long-term investment in the site in the hope of eventual financial or other rewards.

Thus, whilst there is insufficient evidence to conclude that SteemIt is successfully encouraging users to generate useful content, this may change in the future.

From a wider perspective, if the SteemIt model works then it may be followed by other sites introducing similar micropayment mechanisms. The extent to which it can be imitated may also depend on the demand for cryptocurrency or the success of advertising, since the site's model relies on increasing demand from investors or sufficient advertising revenue. Whilst the investment only approach has been successful so far for BitCoin, each new cryptocurrency increases the supply and there will presumably be a point at which the market for cryptocurrencies is saturated and new coins are not needed and so advertising money seems to be the only long-term solution.

References

1. Larimer D, Scott N, Zavgorodnev V, Johnson B, Calfee J, and Vandeberg M. Steem: An incentivized, blockchain-based social media platform. https://steem.io/SteemWhitePaper.pdf (2016, accessed 9 November 2017).
2. Bijker WE, Hughes TP, Pinch T, and Douglas DG. (eds). *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: MIT press, 2012.
3. Kling R, McKim G, and King A. A bit more to it: scholarly communication forums as socio-technical interaction networks. *Journal of the Association for Information Science and Technology* 2003; 54: 47-67.
4. Meyer E. Socio-technical interaction networks: A discussion of the strengths, weaknesses and future of Kling's STIN model. In: Berleur J. Nurminen MI, Impagliazzo J. (eds) *Social informatics: an information society for all? In remembrance of Rob Kling*. Berlin, Germany: Springer, 2006, pp. 37-48.
5. Massanari A. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 2017; 19: 329-346.
6. Hemp P. Avatar-based marketing. *Harvard Business Review* 2006; 84: 48-57.
7. 99Bitcoins. Making money with SteemIt (Steem Dollars and Steem Power explained). https://99bitcoins.com/making-money-with-steemit-steem-dollars-and-steem-power-explained/ (2016, accessed 9 November 2017).
8. SteemIt. FAQ. https://steemit.com/faq.html (2017, accessed 22 July 2017).
9. Ats-david. Advertising revenue and distributions on the Steem blockchain. https://steemit.com/steemit/@ats-david/advertising-revenue-and-distributions-on-the-steem-blockchain (2017, accessed 9 November 2017).
10. SteemItBlog. Introducing promoted content. https://steemit.com/steemit/@steemitblog/introducing-promoted-content (2016, accessed 9 November 2017).
11. Donkeypong. Still confused by Steem, Steem Dollars, and Steem Power? The power plant analogy. https://steemit.com/steemit/@donkeypong/still-confused-by-steem-steem-dollars-and-steem-power-the-power-plant-analogy (2016, accessed 9 November 2017).
12. Boviard C. New Digital Currency Steem Provokes Doubt of Market Observers. http://www.coindesk.com/steem-provokes-doubt-market-observers/ (2016, accessed 9 November 2017).
13. Bartoletti M, Carta S, Cimoli T, and Saia R. Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. arXiv preprint 2017 arXiv:1703.03779

14. Grinberg R. Bitcoin: An innovative alternative digital currency. *Hastings Science and Technology Law Journal* 2012; 4: 159-218.
15. Tschorsch F, and Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys & Tutorials* 2016; 18: 2084-2123.
16. Ciaian P, Rajcaniova M, and Kancs DA. The economics of BitCoin price formation. *Applied Economics* 2016; 48: 1799-1815.
17. Wan Z, Lo D, Xia X, and Cai L. Bug characteristics in blockchain systems: a large-scale empirical study. In: *Proceedings of the 14th International Conference on Mining Software Repositories*. Los Alamitos: IEEE Press, 2017 pp. 413-424.
18. Parks MR, and Floyd K. Making friends in cyberspace. *Journal of Computer-Mediated Communication* 1996; 1: doi:10.1111/j.1083-6101.1996.tb00176.x
19. Smith MS, and Giraud-Carrier C. Bonding vs. bridging social capital: A case study in Twitter. In: *IEEE Second International Conference on Social Computing (SocialCom 2010)*. Los Alamitos, CA: IEEE Press, 2010, pp. 385-392.
20. Bohn A, Buchta C, Hornik K, and Mair P. Making friends and communicating on Facebook: Implications for the access to social capital. *Social Networks* 2014; 37: 29-41.
21. Ellison NB, Steinfield C, and Lampe C. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society* 2011; 13: 873-892.
22. Ransbotham S, and Kane GC. Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia. *MIS Quarterly* 2011; 35: 613-627.
23. Cole H, and Griffiths MD. Social interactions in massively multiplayer online role-playing gamers. *CyberPsychology & Behavior* 2007; 10: 575-583.
24. Mislove A, Viswanath B, Gummadi KP, and Druschel P. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*. New York, NY: ACM Press, 2010, pp. 251-260.
25. Thelwall M. Homophily in MySpace, *Journal of the American Society for Information Science and Technology* 2009; 60: 219-231.
26. Benevenuto F, Rodrigues T, Cha M, and Almeida V. Characterizing user behavior in online social networks. In: *Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference*. New York, NY: ACM Press, 2009, pp. 49-62.
27. Wen S, Jiang J, Xiang Y, Yu S, and Zhou W. Are the popular users always important for information dissemination in online social networks? *IEEE Network* 2014; 28: 64-67.
28. Subramani MR, and Rajagopalan B. Knowledge-sharing and influence in online social networks via viral marketing. *Communications of the ACM* 2003; 46: 300-307.
29. Wilkinson D, and Thelwall M. Social network site changes over time: The case of MySpace. *Journal of the American Society for Information Science and Technology* 2010; 61: 2311-2323.
30. Rogers EM. *Diffusion of innovations*. New York, NY: Simon and Schuster, 2010.
31. Putnam RD. Bowling alone: America's declining social capital. *Journal of Democracy* 1995; 6: 65-78.

32. Chiu CM, Hsu MH, and Wang ET. Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems* 2006; 42: 1872-1888.
33. Glenski M, Pennycuff C, and Weninger T. Consumers and curators: Browsing and voting patterns on Reddit. *IEEE Transactions on Computational Social Systems* 2017; in press.
34. Chung JY, and Buhalis D. Information needs in online social networks. *Information Technology & Tourism* 2008; 10: 267-281.
35. Kim S, and Oh S. Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the Association for Information Science and Technology* 2009; 60: 716-727.
36. Gazan R. Specialists and synthesists in a question answering community. *Proceedings of the Association for Information Science and Technology* 2006; 43: 1-10.
37. Frey BS, and Jegen R. Motivation crowding theory. *Journal of Economic Surveys* 2001; 15: 589-611.
38. Liu CC, Liang TP, Rajagopalan, B, and Sambamurthy, V. The crowding effect of rewards on knowledge-sharing behavior in virtual communities. In: *Pacific Asia Conference on Information Systems (PACIS 2011)* 2011 pp. 116-129.
39. Cui J, Wang L, Feng H, and Teng Y. Empirical study of the motivations of e-wom spreading on online feedback System in China. In: *Pacific Asia Conference on Information Systems (PACIS 2014)* 2014, pp. 251-261.
40. Duggan M, and Smith A. 6% of online adults are Reddit users. *Pew Internet & American Life Project*. http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_reddit_usage_2013.pdf_2013 (accessed 9 November 2017).
41. Ovadia S. More than just cat pictures: Reddit as a curated news source. *Behavioral & Social Sciences Librarian* 2015; 34: 37-40.
42. Gilbert E. Widespread underprovision on Reddit. In: *Proceedings of the 2013 conference on Computer supported cooperative work*. New York, NY: ACM Press, 2013, pp. 803-808.
43. Matias JN. Going dark: Social factors in collective action against platform operators in the Reddit blackout. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI2016)*. New York, NY: ACM Press, 2016, pp. 1138-1151.
44. Singer P, Flöck F, Meinhart C, Zeitfogel E, and Strohmaier M. Evolution of Reddit: from the front page of the internet to a self-referential community? In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY: ACM Press, 2014, pp. 517-522.
45. Suran M, and Kilgo DK. Freedom from the press? How anonymous gatekeepers on Reddit covered the Boston Marathon bombing. *Journalism Studies* 2017; 18: 1035-1051.
46. Dosono B, Semaan B, and Hemsley J. Exploring AAPI identity online: Political ideology as a factor affecting identity work on Reddit. In: *Proceedings of the 2017 Conference on Human Factors in Computing Systems (CHI2017)*. New York, NY: ACM Press, 2017, pp. 2528-2535.
47. Mills RA. Pop-up political advocacy communities on Reddit.com: SandersForPresident and The Donald. *AI & Society* 2017, 1-16.

48. Gibeault MJ. Embracing Geek culture in undergraduate library instruction: The TIL subreddit for resource evaluation and qualitative assessment. *The Reference Librarian* 2016; 57: 205-212.

49. Singer P, Ferrara E, Kooti F, Strohmaier M, and Lerman K. Evidence of online performance deterioration in user sessions on Reddit. *PloS One* 2016; 11: e0161636.

50. Jamali S, and Rangwala H. Digging Digg: Comment mining, popularity prediction, and social network analysis. In: *International Conference on Web Information Systems and Mining (WISM 2009)*. Los Alamitos: IEEE Press, 2009, pp. 32-38.

51. Pavlopoulou MEG, Tzortzis G, Vogiatzis D, and Paliouras G. Predicting the evolution of communities in social networks using structural and temporal features. In: *12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP2017)*. Los Alamitos, CA: IEEE Press, 2017, pp. 40-45.

52. Wang S, Chen TH, and Hassan AE Understanding the factors for fast answers in technical Q&A websites. *Empirical Software Engineering* 2017, in press; 1-42. doi: 10.1007/s10664-017-9558-5

53. Bornfeld B, and Rafaeli S. Gamifying with badges: A big data natural experiment on Stack Exchange. *First Monday* 2017, 22 https://firstmonday.org/ojs/index.php/fm/article/view/7299/6301

54. Burghardt K, Alsina EF, Girvan M, Rand W, and Lerman K. The myopia of crowds: Cognitive load and collective evaluation of answers on Stack Exchange. *PloS One* 2017; 12: e0173610.

55. Gao S, Ma J, and Chen Z. Effective and effortless features for popularity prediction in microblogging network. In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY: ACM Press, 2014, pp. 269-270.

56. Suh B, Hong L, Pirolli P, and Chi EH. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In: *IEEE second international conference on Social computing (socialcom2010)*. Los Alamitos, CA: IEEE Press, 2010, pp. 177-184.

57. Chen J, Liu Y, and Zou M. User emotion for modeling retweeting behaviors. *Neural Networks* 2017; 96: 11-21.

58. Mahdavi M, Asadpour M, and Ghavami SM. A comprehensive analysis of tweet content and its impact on popularity. In: *8th International Symposium on Telecommunications (IST2016)*. Los Alamitos, CA: IEEE Press, 2016, pp. 559-564.

59. Kupavskii A, Ostroumova L, Umnov A, Usachev S, Serdyukov P, Gusev G, and Kustarev A. Prediction of retweet cascade size over time. In: *Proceedings of the 21st ACM international conference on Information and knowledge management*. New York, NY: ACM Press, 2012, pp. 2335-2338.

60. Petrovic S, Osborne M, and Lavrenko V. (2011). RT to Win! Predicting Message Propagation in Twitter. *ICWSM2011*. Los Alamitos, LA: IEEE Press (pp. 586-589).

61. Comarela G, Crovella M, Almeida V, and Benevenuto F. Understanding factors that affect response rates in Twitter. In: *Proceedings of the 23rd ACM conference on Hypertext and social media*. New York, NY: ACM Press, 2012, pp. 123-132.

62. Wu B, and Shen H. Analyzing and predicting news popularity on Twitter. *International Journal of Information Management* 2015; 35: 702-711.

63. Costa AF, Traina AJM, Traina C, and Faloutsos C. Vote-and-comment: modeling the coevolution of user interactions in social voting web sites. In: *IEEE 16th International*

Conference on Data Mining (ICDM2016). Los Alamitos, CA: IEEE Press, 2016, pp. 91-100.

64. Glenski M, and Weninger T. Rating effects on social news posts and comments. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2017; 8: 78.

65. Leavitt A, and Robinson JJ. The role of information visibility in network gatekeeping: Information aggregation on Reddit during crisis events. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW2017)*. New York, NY: ACM Press, 2017, pp. 1246-1261.

66. Arcang. [STEEMSQL.COM] How to create a Steem analytic report with Microsoft Excel. https://steemit.com/steemit/@arcange/steemsql-com-how-to-create-a-steem-analytic-report-with-microsoft-excel (2016, accessed 9 November 2017).

67. Thelwall M, Buckley K, Paltoglou G. Cai D, and Kappas A. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 2010; 61: 2544–2558.

68. Taboada M, Brooke J, Tofiloski M, Voll K, and Stede M. Lexicon-based methods for sentiment analysis. *Computational Linguistics* 2011; 37: 267-307.

69. Thelwall M, Buckley K, and Paltoglou G. Sentiment strength detection for the social Web, *Journal of the American Society for Information Science and Technology* 2012; 63: 163-173.

70. Berk RA, Western B, and Weiss RE. Statistical inference for apparent populations. *Sociological Methodology* 1995; 25: 421-458.

71. Wallach HM. Topic modeling: beyond bag-of-words. In: *Proceedings of the 23rd international conference on Machine learning*. New York, NY: ACM Press, 2006, pp. 977-984.

72. Beijnon B, Ha T, Kim S, and Kim JH. Examining user perceptions of smartwatch through dynamic topic modeling. *Telematics and Informatics*, in press.

73. Liu Y, and Li H. Exploring the impact of use context on mobile hedonic services adoption: An empirical study on mobile gaming in China. *Computers in Human Behavior* 2011; 27: 890-898.

74. Bourdieu P. The forms of capital (1986). In: Szeman I, Kaposy T. (eds) *Cultural theory: An anthology*. New York, NY: Wiley-Blackwell, 2011, pp. 81-93.