ScamCoins, S*** Posters, and the Search for the Next Bitcoin™: Collective Sensemaking in Cryptocurrency Discussions

EAMAN JAHANI, MIT IDSS, USA
PETER M. KRAFFT, MIT CSAIL, USA
YOSHIHIKO SUHARA, MIT Media Lab, USA
ESTEBAN MORO, MIT Media Lab, USA and GISC & Department of Mathematics, Universidad Carlos III de Madrid, Spain
ALEX 'SANDY' PENTLAND, MIT Media Lab, USA

Participants in cryptocurrency markets are in constant communication with each other about the latest coins and news releases. Do these conversations build hype through the contagiousness of excitement, help the community process information, or play some other role? Using a novel dataset from a major cryptocurrency forum, we conduct an exploratory study of the characteristics of online discussion around cryptocurrencies. Through a regression analysis, we find that coins with more information available and higher levels of technical innovation are associated with higher quality discussion. People who talk about "serious" coins tend to participate in discussion displaying signatures of collective intelligence and information processing, while people who talk about "less serious" coins tend to display signatures of hype and naïvety. Interviews with experienced forum members also confirm these quantitative findings. These results highlight the varied roles of discussion in the cryptocurrency ecosystem and suggest that discussion of serious coins may be oriented towards earnest, perhaps more accurate, attempts at discovering which coins are likely to succeed.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; *Collaborative and social computing design and evaluation methods*; • **Social and professional topics** → *User characteristics*; • **Applied computing** → Law, social and behavioral sciences;

Keywords: Cryptocurrency; Online Discussion; Collective Intelligence; Information Processing; Computational Social Science

ACM Reference Format:
Eaman Jahani, Peter M. Krafft, Yoshihiko Suhara, Esteban Moro, and Alex 'Sandy' Pentland. 2018. ScamCoins, S*** Posters, and the Search for the Next Bitcoin™: Collective Sensemaking in Cryptocurrency Discussions. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 79 (November 2018). ACM, New York, NY. 28 pages. https://doi.org/10.1145/3274348

Authors' addresses: Eaman Jahani, MIT IDSS, 50 Ames St, Cambridge, MA, 02142, USA, eaman@mit.edu; Peter M. Krafft, MIT CSAIL, 77 Massachusetts Avenue, Cambridge, 02139, MA, USA, pkrafft@csail.mit.edu; Yoshihiko Suhara, MIT Media Lab, 20 Ames St, Cambridge, MA, 02142, USA, suhara@mit.edu; Esteban Moro, MIT Media Lab, 20 Ames St, Cambridge, MA, 02142, USA , GISC & Department of Mathematics, Universidad Carlos III de Madrid, Spain, emoro@mit.edu; Alex 'Sandy' Pentland, MIT Media Lab, 20 Ames St, Cambridge, MA, 02142, USA, sandy@media.mit.edu.

2573-0142/2018/11-ART79 $15.00
https://doi.org/10.1145/3274348

1 INTRODUCTION

In the past years, cryptocurrency markets have gone from relatively obscure to being, at least for a time, a massive global economic force with huge total market value and widespread societal impact [45]. Yet the cryptocurrency ecosystem is plagued by uncertainty. The uncertainty of the cryptocurrency ecosystem is exacerbated by the large array of cryptocurrency variants, alternative coins or altcoins, that are introduced to the market on a daily basis [43, 55]. One important question in this ecosystem is whether cryptocurrencies as a class of assets or as a class of algorithms will stand the test of time: Are cryptocurrencies merely a temporary investment hype, or do they offer genuine technological innovation that facilitates transactions in the digital economy? And if cryptocurrencies do end up overall having sticking power, which cryptocurrencies will become dominant in the long-term?

Active participants in cryptocurrency markets come face-to-face with these questions every day. These participants must attempt to make sense of the information available about the catalog of existing cryptocurrencies in order to make their guesses as educated as possible. Not an easy task. Making sense of the information available can be difficult because of the technical skills required to understand the nuances that distinguish cryptocurrencies, and because information about particular cryptocurrencies can be sparse and distributed. How well do cryptocurrency users distinguish hype from fundamental value? To what extent can they determine the value for themselves versus having to collaborate with others to determine value? How successful is the collective effort of the cryptocurrency community at discovering the altcoins that have real technical merits?

In this paper, we investigate to what extent the cryptocurrency community discussions are building excessive hype over and above reasonable interpretations of public information, and to what extent the community is in the process of determining the true value by communicating and processing available information. To answer this question, we leverage the fact that *truth-seeking discussions produce collective opinions that are sensibly derived from and consequently constrained by the available information, whereas hype-based discussions lead to a level of excitement that is in principle uncorrelated with the amount of information available about the coin.*

Both of these discussion patterns in the crypto community are possible [11]. There have certainly been some examples of bubbles bursting (e.g., the Auroracoin crash in 2014), but some have even argued that the whole crypto-ecosystem is a bubble, i.e. primarily hype-based, since its fundamental value is zero [7]. In contrast, many have pointed to the remarkable potential of the cryptocurrencies to replace fiat money altogether and unleash new technologies [57]. The true picture is likely to be somewhere in between these two views. While it might be true that cryptocurrency markets, and in particular Bitcoin, were inefficient and hype-based, recent evidence suggests the cryptocurrency market is moving towards efficiency [56]. Recent work [19] found mixed results on the extent to which cryptocurrency community is truth-seeking. Through in-depth interviews, researchers concluded that the community is "split between people who speculate and those who believe in its long-term potential" further noting that fragmentation of cryptocurrency user pool warrants further study [19]. Our work is one step toward this detailed characterization of the community aiming to identify when and how the community is genuinely engaged in technical evaluation of new altcoins.

In order to accomplish this goal, we identify several indicators of "collective sensemaking" processes that indirectly measure the degree of truth-seeking in the community discussions. We appeal to the literature on collective intelligence, which has identified quantitative metrics for discussion quality that are correlated with effective collective problem-solving in groups [14, 24, 32, 59]. Many studies have investigated the fundamentals for effective face-to-face teamwork [22] and identified key signatures of collective intelligence in groups such as proportion of women [59] and equal conversational turn-taking [29, 58]. A few very recent studies have verified that similar signatures of collective intelligence also apply to online distributed communities in a fast-changing environment, an important focus of the CSCW community [16, 17, 32]. Since cryptocurrency users collaborate and communicate in exactly such an online distributed community, we use these metrics in order to assess which cryptocurrency discussions tend to be high quality in the ways that might promote truth-seeking. We use conversational turn-taking in discussion threads as one metric, and leverage exposure to various sub-communities as a proxy for diverse experiences, both of which have been shown to be positively associated with collective intelligence [5, 24, 36, 58]. We also rely on seniority as a proxy for individual learning and expertise [53]. Using these indicators, we can measure the extent of collective sensemaking in the discussion of various cryptocurrencies.

Another challenge after measuring these metrics is the lack of an absolute baseline for levels of these indicators that are representative of meaningful collective sensemaking. To address this challenge, we use the fact that there is great variation in the level of *objective* information available to process by the community about different altcoins. We operationalize information availability with two metrics: (1) *Price Volatility*: Although price fluctuations are influenced by liquidity, market or macro-economic uncertainty, the finance literature indicates that one of the main sources of price volatility is related to the specific uncertainty about an asset [3, 12, 13]. In this view, lower volatility corresponds to lower uncertainty which is directly affected by the level of public information available about an asset. (2) *Coin Technicality*: Non-triviality of a cryptocurrency or an objective measure of its technical innovation serves as another operationalization of information availability. For such coins, the community has a lot more information on technical advancements of the cryptocurrency, which are typically published as white papers, to discuss and make sense of.

If the cryptocurrency ecosystem is purely hype-based, we would expect little collective sensemaking to occur in the community discussions, and importantly, no statistically significant difference between the discussions of coins with differing levels of information available about them. We can evaluate this claim by checking whether there is a significant difference in indicators of collective sensemaking between discussion of cryptocurrencies with more or less information available. Our results suggest otherwise. We observe more collective sensemaking in discussion of more technical (less volatile) than less technical (more volatile) cryptocurrencies. We reach this conclusion through an exploratory study of a novel dataset from the predominant online forum dedicated to cryptocurrencies. Using a regression analysis, we find that discussion metrics associated with indicators of high collective intelligence are correlated with lower volatility and more substantive coins. Coins with less healthy discussion pages, with lower measures of collective intelligence, have higher market price volatility and less substantial technical innovations associated with them. The lower volatility, more substantive coins that display more indicators of high-quality discussion also on average have older accounts—presumably associated with more experienced users. To qualitatively evaluate these quantitative findings, we conducted brief interviews with a few senior forum members and asked them about the discussion differences between technical and non-technical altcoins. These interviews confirm our findings from the regression analysis.

These results suggest that discussion networks in this context have varied but somewhat predictable functions. With “more serious” coins—for which there is more information available or more inherent innovation—discussion seems to serve more of a truth-seeking role, perhaps in an attempt to distinguish which coins among plausible contenders are most likely to succeed. For “less serious” coins—for which there is less information available or less inherent innovation—discussion may be more hype-based. These results further refine previous findings [19] that there are indeed two sub-communities within the larger cryptocurrency user base. Supporting this interpretation, we see that a simple cluster analysis of the online discussions reveals that the userbases of high and low volatility coins are almost separate. One subcommunity is excited about the socio-technical potential of cryptocurrencies and is well-versed in mechanics of various cryptocurrency protocols (i.e. they see Bitcoin and its variants as a currency rather than an investment). The second sub-community is engaged in speculation and less committed to the ecosystem technical innovations. Our claims are further strengthened when we consider recent findings that suggest cryptocurrency volatility is a good indicator of coin quality, having a significant long-run impact on the future of the cryptocurrency [52].

2 RELATED WORK

In this work, we investigate the interaction of structural[1] determinants of collective intelligence in online discussions with technicality and information availability about the digital currencies that manifest themselves in the price stability. There are four main factors that are associated with price movements of digital currencies: (1) Supply and demand (2) Legal issues on adoption (3) Macro-economic factors such as interest rate or stock market (4) Speculation and attractiveness of the digital currency based on its potential [47]. Recent research on cryptocurrency market, while diverse, addresses one of the above factors and its relationship with prices or trade volume. A wide array of recent studies have looked at the economics of digital currencies (items 1 and 3 above). An objective pricing model for value formation of crypto coins based on supply rate, mining difficulty and the competition among producers is developed in [23]. Another study relates Bitcoin price to the price of gold, hash rate and output volume [6]. The quality of supply side through developer activity and its relationship with market growth is evaluated in [43]. A similar study [9] proposes an equilibrium model for Bitcoin which evaluates the welfare implication of several supply parameters in its design. In contrast to these recent works, our study is a descriptive attempt in unpacking the last factor (item 4 above) which solely depends on the information processing by the community of actors in the cryptocurrency market.

There is an extensive literature on the prediction of stock market prices using social media [4, 8, 54]. However, studies in the context of digital currencies are still very scant. A few recent papers have investigated the relationship between cryptocurrency prices and the social factors in the cryptomarket (e.g. market sentiment). Most works have focused on Bitcoin or a few altcoins with the largest volume. The work by [20] uses aggregate social insights such as word-of-mouth volume and valence on social media to design an optimal Bitcoin trading strategy. Another work focuses on Bitcoin and the top four altcoins (Monero, Dash, Ethereum, and Litcoin) and finds evidence of the long-run effect of social attractiveness measured in terms of Google search term frequency on prices [52]. Importantly, their findings indicate that social attention and attractiveness of these 5 cryptocurrencies take effect on their value formation only over long-term which suggests that at least for these 5 cryptocurrencies, social attention is associated with their underlying quality, rather than a short-term market hype. Several other studies, all focusing on Bitcoin, looked at social media such as Twitter [27, 39] and search engine trends [37, 46, 60] to distill market sentiment and predict future price fluctuations. The main conclusion from these studies is that social attention and opinions about different assets, specifically in relation to major shocks, can predict future trading and thus inform us about the evolution of the cryptocurrency market.

The studies mentioned above [20, 27, 37, 39, 52, 60] were successful in predicting price fluctuations, but did not provide any insights on the nature of these fluctuations in relation to collective information processing by active traders. Furthermore, as the goal was mainly prediction, they analyzed the social media from the general public which is not guaranteed to be from the active

[1]We refer to our indicators as structural since we derive them solely based on the structure and not the content of the discussion.

traders or engaged community members. For this reason, we focus on discussions among cryptocurrency users in their main and the oldest online forum community (https://bitcointalk.org). A large number of discussion threads on the forum revolve around successful trading which suggests a significant fraction of active users on the forum are traders with skin in the game.

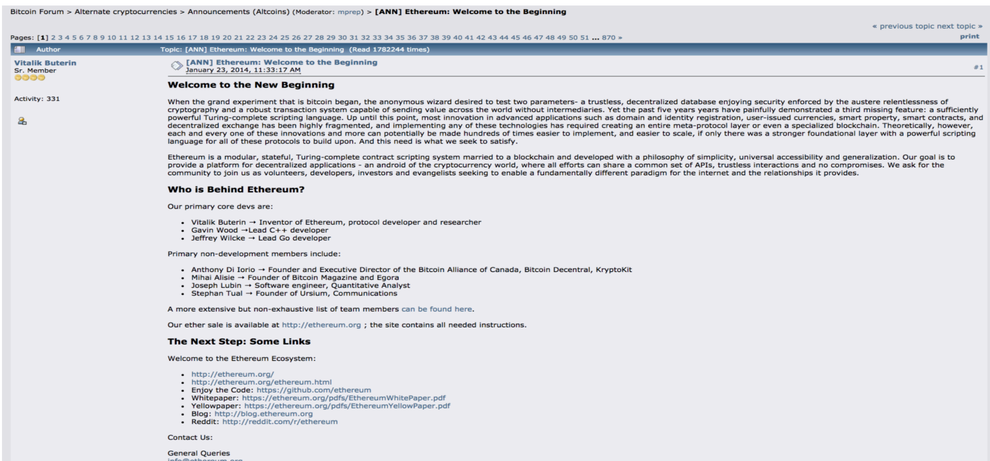
A few recent studies have relied on the discussion posts in https://bitcointalk.org online forum to investigate a diverse set of questions, not all related to the price movements. One study examined how the evolution of infrastructure code can lead to the creation of new sub-organizations in distributed digital communities such as crypto community [1]. By generating a discourse time series in https://bitcointalk.org forum using LDA topic modeling, [1] studied self-organizing patterns in Bitcoin community after important code forks in Bitcoin repository. A similar work detected new trends in crypto community using topic modeling and validated its predictive power by comparing to major events such as fraudulent schemes and economic concerns [38]. The work of [51] argues that in addition to discovering value, collective sensemaking in discussion forums may be a part of the process through which collective valuation of assets is endogenously created. Perhaps the most similar to ours is a series of works by Kim et al. which aim to predict Bitcoin price fluctuation by analyzing the content of posts made in the forum through topic modeling [31] or sentiment analysis [30]. While our study uses the same data sources, namely forum posts and price time series, our goals are different. First, our goal is not price prediction, but rather discovering the difference between collective information processes of crypto coins with high and low levels of available information (and quality). Second, in addition to analyzing the content of posts, we focus on structural patterns of discussion such as the age of users or exposure to the larger community.

Another contribution which separates this study from previous work is our focus on a large collection of altcoins, not limited to the most successful cryptocurrencies with the largest volume, since our aim is to understand social factors that separate coins with varying levels of uncertainty around their technical value, no matter if they are successful or not. Some recent works have studied the price dynamics of a large number of altcoins. Researchers have recently demonstrated through an online field experiment that cryptocurrency market dynamics may be highly susceptible to peer influence effects [35]. In this experiment, researchers implemented bots to trade in hundreds of cryptocurrency markets and showed that other traders in these markets were more likely to buy after the bots bought. These results suggest a degree of faddishness in cryptocurrency markets, but the measurable effects in this experiment were short-lived. A similar work [18] has observationally documented the susceptibility of cryptocurrency markets to manipulation. A recent work by [15] looked at the price history of all altcoins and discovered several stable statistical properties of the whole market that match the system growth expectations from an evolutionary model. While these studies have provided insights into the general dynamics of the cryptocurrency trading prices, they have not explored how they might be associated with the discussion and information sharing within the cryptocurrency community.

3 DEFINITIONS

3.1 Hype-Based Discussion

We define hype-based discussion as discussion that is not constrained by available information. Hype-based discussion is an endogenous social process in which excitement or pessimism emerges from people reflecting their opinions off each other. Sociologists have argued that success and failure in cultural markets are driven in part by hype-based social processes [49]. These social feedback processes may include self-reinforcing emotions of excitement or worry [21] or behavioral imitative contagion process [35]. In the context of cryptocurrencies, hype-based social processes


Bitcoin Forum > Alternate cryptocurrencies > Announcements (Altcoins) (Moderator: mprep) > [ANN] Ethereum: Welcome to the Beginning

« previous topic next topic »

Pages: [1] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 ... 870 »

print

Author | Topic: [ANN] Ethereum: Welcome to the Beginning (Read 1782244 times)

Vitalik Buterin
Sr. Member

Activity: 331

[ANN] Ethereum: Welcome to the Beginning
January 23, 2014, 11:33:17 AM

#1

Welcome to the New Beginning

When the grand experiment that is bitcoin began, the anonymous wizard desired to test two parameters- a trustless, decentralized database enjoying security enforced by the austere relentlessness of cryptography and a robust transaction system capable of sending value across the world without intermediaries. Yet the past five years years have painfully demonstrated a third missing feature: a sufficiently powerful Turing-complete scripting language. Up until this point, most innovation in advanced applications such as domain and identity registration, user-issued currencies, smart property, smart contracts, and decentralized exchange has been highly fragmented, and implementing any of these technologies has required creating an entire meta-protocol layer or even a specialized blockchain. Theoretically, however, each and every one of these innovations and more can potentially be made hundreds of times easier to implement, and easier to scale, if only there was a stronger foundational layer with a powerful scripting language for all of these protocols to build upon. And this need is what we seek to satisfy.

Ethereum is a modular, stateful, Turing-complete contract scripting system married to a blockchain and developed with a philosophy of simplicity, universal accessibility and generalization. Our goal is to provide a platform for decentralized applications - an android of the cryptocurrency world, where all efforts can share a common set of APIs, trustless interactions and no compromises. We ask for the community to join us as volunteers, developers, investors and evangelists seeking to enable a fundamentally different paradigm for the internet and the relationships it provides.

Who is Behind Ethereum?

Our primary core devs are:

- Vitalik Buterin → Inventor of Ethereum, protocol developer and researcher
- Gavin Wood →Lead C++ developer
- Jeffrey Wilcke → Lead Go developer

Primary non-development members include:

- Anthony Di Iorio → Founder and Executive Director of the Bitcoin Alliance of Canada, Bitcoin Decentral, KryptoKit
- Mihai Alisie → Founder of Bitcoin Magazine and Egora
- Joseph Lubin → Software engineer, Quantitative Analyst
- Stephan Tual → Founder of Ursium, Communications

A more extensive but non-exhaustive list of team members can be found here.

Our ether sale is available at http://ethereum.org ; the site contains all needed instructions.

The Next Step: Some Links

Welcome to the Ethereum Ecosystem:

- http://ethereum.org/
- http://ethereum.org/ethereum.html
- Enjoy the Code: https://github.com/ethereum
- Whitepaper: https://ethereum.org/pdfs/EthereumWhitePaper.pdf
- Yellowpaper: https://ethereum.org/pdfs/EthereumYellowPaper.pdf
- Blog: http://blog.ethereum.org
- Reddit: http://reddit.com/r/ethereum

Contact Us:

General Queries
info@ethereum.org

Fig. 1. The first post of a thread announcing the release of Ethereum for the first time in bitcointalk.org. With 20 replies per page at the time of writing, the thread had received well over 17,400 replies.

would yield random coins becoming popular simply because of bandwagon effects, and hype-based discussion would reflect this excitement unconstrained by information about the coins.

3.2 Truth-Seeking Discussion

We define truth-seeking discussion as discussion involving earnest attempts to make sense of the world. We operationalize truth in this case as the fundamental value of a cryptocurrency. Fundamental value is an influential notion in economics and finance. The fundamental or intrinsic value of an asset is an objective measure of its longtime value which solely depends on its quality in relation to competitors and the market. It is measured as the discounted sum of the asset's cash flow in the future, which reflects the "correct" price of the asset [26]. In the case of a cryptocurrency, the fundamental value would be related to the usefulness of the algorithmic protocol associated with a coin. This operationalization of truth is related to the two metrics we use to measure information availability. Price volatility indicates an uncertainty of fundamental value due to a lack of information being available [3, 12, 13]. The technicality of a coin produces information that should reduce such uncertainty. We suppose that truth-seeking discussions would be oriented towards discovering fundamental value in these terms.

4 DATA DESCRIPTION

We draw upon three sources of data in this study: text and network data from a major online cryptocurrency forum; cryptocurrency price data; and a curated dataset of which coins represent technical innovations as opposed to rebrandings of the Bitcoin protocol or that of another coin with only a few minor changes (e.g., in parameters).

4.1 Online Discussions Forum

We collected all the posts from the most popular cryptocurrency online community, https://bitcointalk.org. This online forum has been around since the early days of Bitcoin in 2009 and has acted as the main venue of discussion around Bitcoin, all altcoins and general cryptocurrency trends and technology within the cryptocurrency community. Once Bitcoin and the earliest altcoins went on the exchange markets in early 2013, bitcointalk.org acted as the main source of information, discussion, and marketing of altcoins among developers and traders.

The main Bitcointalk webpage contains multiple forums, each concentrated on a separate aspect of cryptocurrencies. Each forum consists of many subjects or discussion threads initiated by different users. Our online discussion data consists of the all the posts that were made in all threads between January 2010 and November 2016. During this period, there were more 5.9 million posts made in all forums, of which more than 860,000 posts solely focused on Bitcoin, and more than 4.3 million posts focused on altcoins. We are particularly interested in Altcoins announcement forum which is the most active with more than 3.5 million posts up to November 2016. Community announcement such as exchange clients, the addition of new features and most importantly creation and marketing of new altcoins are announced here. Whenever a new altcoin is ready for launch, the developers usually create a new thread in this forum, discuss its features, provide the necessary information on how to mine the currency and trade it on online exchanges. Consequently, the users participate in the discussion of the coin's merits in the same announcement thread and in some cases speculate about its potential prices in the future. Figure 1 shows an example of the first post in a thread which introduced Ethereum coin, currently with the largest daily volume among all altcoins, for the first time in the Announcement forum.

Each thread in a forum contains several posts or replies, with an average of 10 posts per thread. On the other hand, announcement threads are more active with an average of 556 posts. The reply structure within each thread constitutes the basis of our forum analysis. We manually searched the forums for the announcement thread of each coin for which we also had pricing data. We were able to match 855 coins to their respective announcement thread.

The community had only 10,000 unique users until early 2013; however, it grew considerably faster after 2013 and reached about 85,000 by early 2015 and 115,000 by November 2016. Nevertheless, there are only around 8,500 active users on average within any 30 day period in 2016.

4.2 Price History

We also scraped daily price and volume data for 1,052 altcoins from https://coinmarketcap.com. The historical data spans from the first trading day of each coin on an online exchange, for example, LiteCoin (LTC) starting from April 2013 to November 2016. The daily price[2] and volume data are aggregated over all exchanges that trade an altcoin between 00:00–23:59 UTC on that day. While there were more than 1,000 online exchanges that traded at least one cryptocurrency, only a few such as Bitfinex, BTC-E, and Cryptsy carried the bulk of transaction volume during the period of our data.

4.3 Technical Coins

Finally, we use the data released in association with recent work [35] which included a binary indicator whether each coin offered any technological innovation over the existing cryptocurrencies at the time of its launch. As the intersection of [35] data with our coins had only 10 technical coins, we manually added 16 technical coins to their data to increase the power of our tests. We did this by first adding a few innovative coins which have introduced wholly new ideas to the cryptocurrency ecosystem, have gained wide-spread success ever since their launch and are missing in [35] (e.g., Ethereum and Monero). Second, we augmented the data by reviewing which coins are promoted as innovative by a large number of online articles and altcoin forum users. Although this

[2]The price is expressed in US dollars.

collection procedure may be subject to selection bias, we gain confidence in our results from the correspondence between the measurements we get from price and the volatility measurements we get from the technicality.

5 METHODS

The general strategy of our approach is to define several metrics that ultimately come into play in a regression analysis that we use to explore the characteristics of discussion in the cryptocurrency ecosystem. These metrics draw upon the collective intelligence and the finance literature to quantify the quality of discussion and measure the amount of information available about each cryptocurrency. In our regression analysis, we examine how the discussion metrics we define are related to two operationalizations of the amount of information available about the cryptocurrencies, volatility, and technicality.

5.1 Metrics

We refer to the first three variables introduced here as *discussion variables*. The fourth and fifth variables are the two operationalizations of information availability.

(1) **Equal participation by community members:** Existing work has found that highest quality wiki pages are created by a large number of participants with few edits, none of whom contribute disproportionately to the final article [2, 44]. Based on the observation that most contributions in Wikipedia originate from users with a small number of contributions, [33] argues that the essential force behind Wikipedia success is the wisdom of the crowd, a large group of equally participating users. The equal contribution of group members has also been correlated with collective intelligence in laboratory studies [17, 29, 32, 59]. In other contexts, especially for the early promotion of products, a small core and highly active user base which leads the community plays a crucial role in the success of online communities [25, 28, 34, 40], in particular by setting the direction of discussion and providing timely and accurate information. Our first variable captures the extent to which the participants in the thread contribute equally to the discussion of the coin, a pattern similar to turn-taking in face-to-face discussions [29]. An even presence by all the user in the announcement thread indicates a wide and distributed community of the coin users with equal engagement by all the members. On the other hand, high engagement by only a few users likely indicates the presence of a small core community committed to the success of the coin. *Either of these activity patterns could be signs of healthy discussion or sensemaking at different stages of a coin's exposure to the market.*

Once the coin is established and has been present on the market for a long time, equal participation by a broad user base is indicative of endorsement by a wide community. For example, in the 200 days leading to 2016-11-01, less than 15% of 319 posts in Dogecoin (DOGE) announcement thread were made the top two most active contributors. DOGE prices were one of the most stable during the past few years. On the other hand, more than 46% of total 1,382 posts in CrownCoin (CRW) announcement thread were made by only 2 users during the same period. At this level of imbalance, it is hard to imagine that meaningful conversations about the technical aspects of the coin take place with a large community of the users. In fact, CRW experienced one of the highest levels of price volatility in our sample between 2016-08 and 2016-11. These observations are in fact consistent with the previous finding on the success of Wikipedia, as another online collaborative community.

On the other hand, it might be difficult to engender a wide community of enthusiasts about a newly created coin immediately after its launch. During this period, a small community of core users who promote the coin can lead to its success and popularity in the larger community. For example, 4 of the top 6 contributors to NEM coin (XEM) discussion thread prior to its launch were dedicated marketers. Together, they accounted for more than 17% of the 15,299 posts. XEM had a stable price after its launch without any large movements over a short span. In contrast, out of 2,582 posts that were made on the announcement thread of MoonCoin (MOON) prior to its trading, less than 2% were made by the development team. Incidentally, MOON experienced a large price variation in its first 100 days.

The extent to which contributions in an announcement thread are equal can be measured with the Shannon entropy of the number of posts made by each user. In other words, we constructed the empirical distribution of the number of posts per users and measured its normalized entropy:

$$H(N) = \frac{-\sum_{n=1}^{\hat{N}_{max}} \hat{p}_n \log(\hat{p}_n)}{\log(\hat{N}_{max})}, \quad (1)$$

where N corresponds to the (random) number of posts made by each user, $\hat{N}_{max}$ is the observed maximum number of posts by any user and $\hat{p}_n$ is the empirical fraction of users who have made n posts in the announcement thread. As entropy scales with the sample space size, we normalized it using $\log(\hat{N}_{max})$ so that it is always between 0 and 1. The larger the normalized entropy is, the closer this distribution is to the uniform in which all users are equal in terms of contribution.

(2) **Information flow from the rest of the community:** Access to a larger pool of information is closely related to the concept of diversity within groups. It has been shown groups composed of diverse opinions and experience often achieve a higher performance due to their ability to borrow and combine ideas from seemingly unrelated experiences [14, 32, 59]. Being connected to individuals with diverse demographic, cultural and behavioral characteristics may increase one's productivity. This form of diversity is referred to as identity diversity by [24]. They show that in a problem-solving setting, diverse groups (using a variety of perspective and heuristics) may outperform groups made of high-ability problem solvers. *In our context, we can view the community of a crypto coin as a group whose main problem is to discover and evaluate the true novelty of a new altcoin, a collective sensemaking problem.* In this view, a connection to numerous informative discussion threads can be treated as a diversity of past experiences.

The forum https://bitcointalk.org is a very large community with tens of thousands of discussion threads focusing on various aspects of cryptocurrency ecosystem. Each thread holds a valuable piece of information, either technical or financial, which could facilitate the creation or accurate evaluation of innovations. It is reasonable to assume that the community engaged in collective sensemaking of coins with a lot of technical information available have access to a larger portion of this vast body of knowledge, compared to the uncertain coins. Coins whose discussion threads have some sort of a connection to other discussion threads in the larger community should benefit from the advantages of those information sources. Effectively, we can focus on aggregated information at the level of each coin, available to members of its respective community and then operationalize access to information spread across the whole bitcointalk.org community as the degree of each announcement thread in the thread network. In this network, nodes are discussion threads and edges correspond to the existence of at least one user who has co-posted in both threads. In this view, members of an altcoin community are the channels through which information and experience flow from other crypto-currency related discussions to the community.

(3) **Seniority of the discussion users:** We also consider the average age of the contributors in the discussion thread during the analysis time interval. Age of each user is measured as the number of days since their first post in bitcointalk.org. As the crypto-currency community has been growing, there has been increasing number of instances where new users introduce a new altcoin immediately after joining the community. These new altcoins often lack the sophistication of the original altcoins and are simple modifications to the already existing stock of altcoins. In

some cases, newly created altcoins are fraudulent and marketed to the community for the sole purpose of pump-and-dump[3], i.e. “ScamCoins”. In contrast, new altcoins that are introduced to the community by already established members of the community are more likely to be truly innovative. For example, Vitalik Buterin, the main creator of Ethereum (ETH), had a very active presence in bitcointalk.org for more than 2.5 years before introducing ETH to the crypto community in January 2014. As of March 2018, ETH has the largest daily volume among altcoins. As another example, the main contributors to LTC (with 5th largest volume among altcoins as of March 2018) in our study period in 2016 are still the same users who joined the discussion when the coin was first introduced in October 2011. Both of ETH and LTC maintained very low levels of price volatility compared to the rest of the market.

There are two potential reasons why average seniority is associated with more stable prices. First, since more senior contributors are more knowledgeable and experienced, they are more capable of detecting possible flaws or designing innovative features. In other words, seniority is a proxy for expertise and individual learning. [53]. Second, more senior contributors are well-established and have acquired valuable social capital within the community. Therefore, they are more likely to provide objective evaluations and accurate information to the community, as inaccuracy comes as a cost to their status.

(4) **Volatility:** We operationalize the level of information available as price volatility. Volatility is a statistics of price history which measures the dispersion of daily returns of an asset. Assuming that prices reflect the market’s perceived valuation of an asset given current available information, volatility measures uncertainty as the variation in the perceived fundamental value of the asset [3]. It has been shown that some of the volatility of stock prices originates from fluctuations in the level of public information about the asset [3, 12, 13]. The riskier the security is due to incomplete information about its fundamentals, the higher its volatility becomes. This means the price of an asset with high uncertainty can change dramatically over a short period in either direction, as new information about its strengths or weakness come to light. In contrast, the price of well-established assets tends to remain stable and vary at a much smaller rate, as they do not experience large information shocks.

The argument above is particularly true in the case of crypto-currencies as the market is still informal without any established means of information seeking on a new coin. In fact, apart from regulatory shocks that affect all altcoins at the same time, likely a major portion of volatility in altcoin prices is due to uncertainty about their true nature, as is suggested by examining price history of many fraudulent altcoins mainly created for orchestrated pump-and-dump schemes. In our context, volatility is computed as the standard deviation of returns over time:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2}$$

$$V = \sqrt{\frac{\sum_{t=1}^{T}(R_t - \bar{R})^2}{T}} \tag{3}$$

where P_t is the price of the asset in day t, R_t is the returns in day t and V is the volatility of the asset measured over a T days period. Finally, we convert the volatility to log scale as it provides a better linear fit to our discussion variables mentioned above.

(5) **Technicality:** We use the level of technicality available for a small subset of coins as a second measure of information availability[4]. Many altcoins are created by simple modifications to the code

[3]A fraudulent scheme for artificially boosting prices through misleading information. It is usually conducted by a group of traders who sell their shares after prices have sufficiently grown.

[4]Results regarding technicality can also be directly interpreted in terms of coin quality.

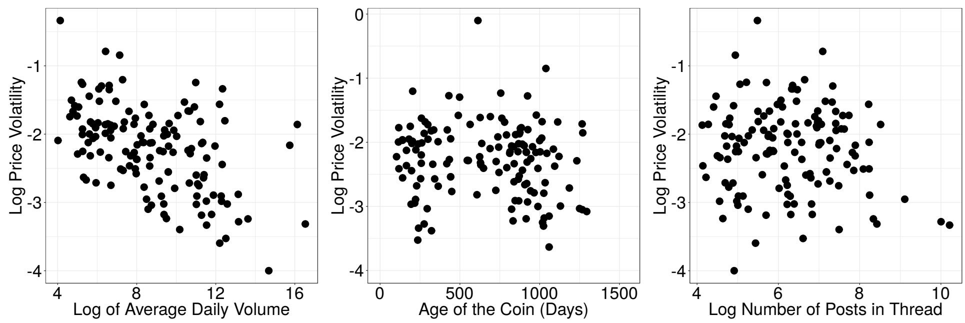


Fig. 2. The relationship between 1) log of average daily volume as the main control variable and volatility (left) and 2) age of the coin as the number of days since it was first announced in bitcointalk.org with its volatility (middle) and 3) log of number of posts made by all users in the announcement thread and volatility (right). There is no significant relationship between volatility and age and the number of posts. In contrast, the relationship between volatility and daily average volume is negative and significant. Volatility and control variables are measured within the 100 or 200 days prior to November 2016 (Design 1 as explained in 5.3).

of already existing coins. These coins differ from their predecessors in only trivial technical aspects, such as the total number of minable coins[5] or the validation protocol of a transaction, which amount to changing a single parameter. Such coins do not possess any technological innovation and as such need to maintain a level of uncertainty about their technical details. In contrast, innovative coins have more technical information available about them. As these coins have substantial innovations, they are often announced along with a detailed white paper or other forms of information on their technical design which can be discussed and made sense of by the community. For this measure, we refer to the work of [35] which produced a binary variable for the technical innovation of a coin through analysis of its GitHub repository, mainly its initial forks.

(6) **Control Variables:** We also consider several other variables to control for endogeneities that might exist in our analysis. For example, the degree of the thread might be confounded with the total level of posting activity in the thread. For this reason, we include the total number of posts made on the announcement thread during the study period as a control variable. We do this to ensure none of our discussion variables (which measure some quality aspect of the discussion) are driven by the mere quantity of discussion activity.

As mentioned before, as the trading volume of an asset decreases its price movements tend to become larger, because any single transaction can constitute a considerable portion of the trading volume and have a drastic effect on the asset price. Therefore, it is essential that we include the average daily volume as another control variable in our analysis. Figure 2 confirms the negative relationship between volume and volatility in our data.

It is also conceivable that earlier altcoins, for example, LTC or DOGE, are more likely to be stable since the community has obtained enough information about them through lengthy discussions and their developers were genuinely concerned about the technological innovations. However, Figure 2 shows that there is no significant relationship between the age of the coin and its volatility in our data set. Nevertheless, we decided to control for the age of the coin in our regression analysis. Figure 3 and Table 1 respectively show the histogram and correlation matrix of discussion and control variables.

[5]A hard limit on the maximum number of digital coins that can ever exist, as determined by the currency protocol.

Table 1. Spearman correlation matrix of Design 1 data. Degree, number of posts and volume are in logs.

	Entropy	Degree	User Age	Num of Posts	Coin Age	Volume
Entropy	1	0.14	0.18	-0.24	0.12	0.37
Degree	0.14	1	-0.26	0.77	-0.26	0.45
User Age	0.18	-0.26	1	-0.39	0.51	-0.01
Num Posts	-0.24	0.77	-0.39	1	-0.36	0.23
Coin Age	0.12	-0.26	0.51	-0.36	1	0.20
Volume	0.37	0.45	-0.01	0.23	0.20	1

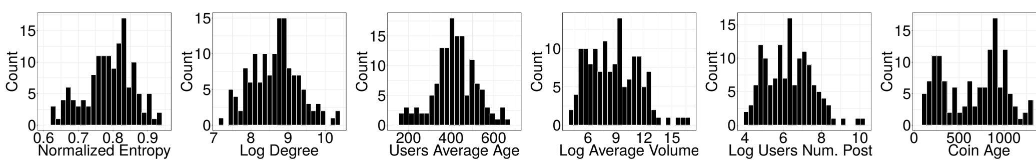


Fig. 3. Histograms of discussion and control variables from Design 1 (as mentioned in 5.3). User age is measured in terms of days and coin age is the number of days since the coin was announced on the forum.

5.2 Data Filtering

We aim to study only those altcoins that have consistent trading and discussion activity over time. The majority of the coins collected from https://coinmarketcap.com do not have enough volume over the study periods to be considered for any meaningful analysis. A considerable portion of altcoins also misses price and volume data for periods that are longer than 10 days (mainly due to zero trading activity in that period). For example, the average daily volume of half the altcoins is less than 15 dollars over a 100 day period ending at 2016-11-01. At such low levels of daily volume, the price movements can be safely characterized as noise, as a single buy/sell request could constitute the whole trading activity of a day, hence prone to rapid price movements. Furthermore, any insight gained from price volatility at such low volumes will be invalid as low volume assets generally exhibit large volatility. As a result, we restrict the data to the set of altcoins whose mean of daily volume over the 100 day analysis period is larger than 50 dollars (this corresponds to a minimum of 20 dollars for 25% quantile of daily volumes or minimum total volume of 40,000 dollars over the 100 days). This leaves us with 377 altcoins with enough volume for any meaningful analysis.

We further need to restrict these altcoins to those which also appear in the online discussion forum and have considerable discussion activity. Limited activity or few posts in an altcoin's announcement thread do not offer a strong signal on the extent of its collective sensemaking. Therefore, we removed any altcoins that had less than 50 posts in their announcement threads during the analysis period. Ideally, we should increase these filtering criteria to get more accurate results on the subset of coins with substantial level of community engagement; however, this would pose two potential problems. First, it could remove data points which exhibit characteristics of hype-based discussions, the very coins we would like to study. Second, it would reduce the size of our data and consequently the power of our regression analysis. Since we are controlling for thread and volume activity in our analysis, we doubt that the signals that our discussion variables capture are merely confounded. Nevertheless, we provide a robustness analysis of these filtering criteria in the Appendix A.1.

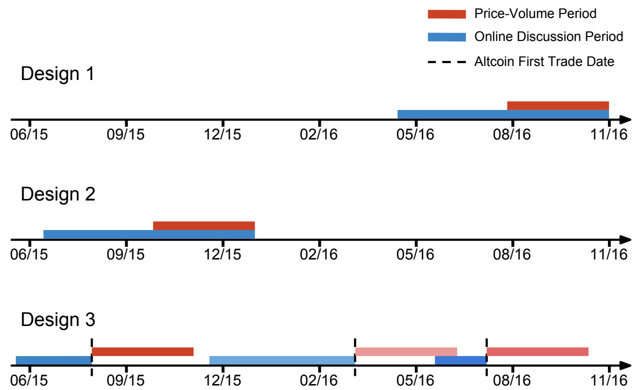
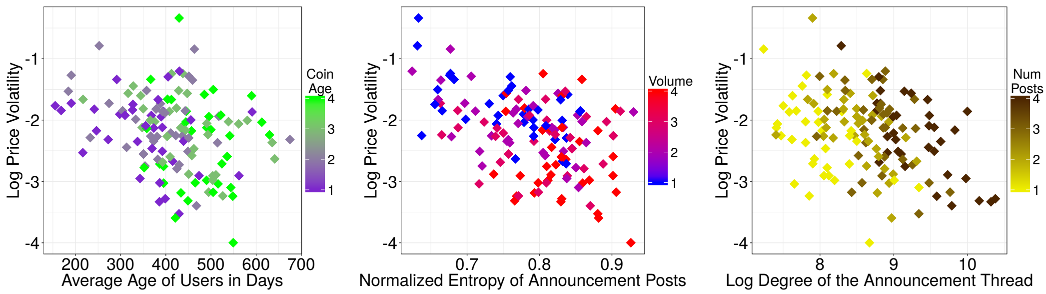


Fig. 4. The periods during which we study market activity and online discussion patterns. Red represents the market price-volume analysis period and blue represent the discussion analysis period. Different colors in Design 3 correspond to different altcoins that start trading at different times.

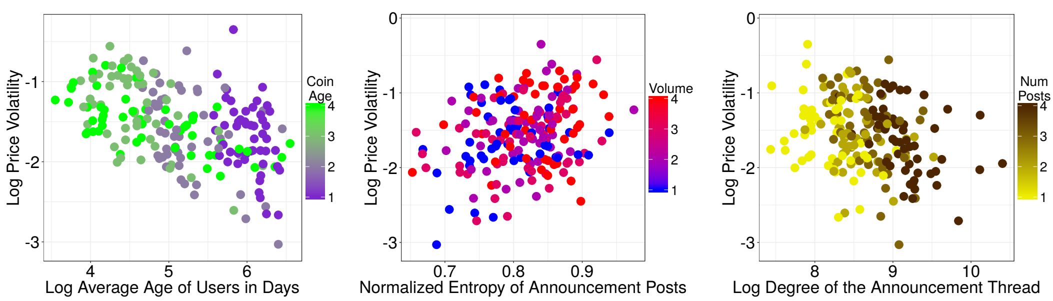
5.3 Analysis Methods

To ensure the robustness of the analysis, we employed three different designs (Designs 1, 2, and 3) that used different periods for extracting online discussions and price-volume information. Figure 4 illustrates the periods during which we measure price-volume dynamics and discussion patterns for each design. In Design 1, we measure price metrics for the 100 days and online discussion for the 200 days before November 2016. As the number of altcoins increased over time, we use the latest time period available in our data (November 2016) so that our data becomes as large as possible and the analysis achieves higher power. To ensure that our results from Design 1 are robust and not due to spurious correlations, we replicate Design 1 and shift the end date of price and discussion periods to January 2016 in Design 2. We note that there must not be any overlap between discussion periods in Designs 1 and 2. We chose a longer period for discussion activity than price patterns for two reasons. First, collective information processing is a lengthy process, and it usually takes a long time for information about an altcoin to appear in community discussion. Second, our online discussion metrics are agnostic to the content of the communication and idiosyncrasy of each community. Such general metrics require long enough period to evaluate the health and substance of the community discussion. In contrast, price and volume are much more responsive to the perceived quality of an asset. We did not want to increase the discussion and price period too much to avoid mixing information shocks and quality signals from the distant past.

In both Designs 1 and 2, volatility and collective sensemaking patterns are endogenous, as one could affect the other and vice-versa, due to their overlapping periods. This would pose a concern to inference if price and discussion were to intensify each other and develop jointly through a feedback loop. In order to limit the extent of this endogeneity, we employ Design 3 in which there is no overlap between price and discussion periods. For each coin, we analyze all discussion activity before the coin starts trading, and evaluate its price patterns during 100 days after its first trade date. Design 3 is more challenging than the first two designs for many reasons. First, there is a distinct time separation between discussion and price periods. This limits the extent to which external and temporary shocks, such as the arrival of new information, could affect both simultaneously. Second, since different coins are introduced to the market at different times, they could experience different levels of discussion and market fixed effects, such as general market-wide volatility. Third, the length of the discussion period varies for different coins, since many altcoin developers initiate



(a) Design 1



(b) Design 3

Fig. 5. The price volatility of the coins over a 100 day period in (a) Design 1 (top row) and (b) Design 3 (bottom row) versus discussion variables: log average age of the users in announcement thread in terms of number of days (left column), normalized entropy of numbers of posts made by each user in the announcement thread (middle column) and the log degree of the announcement thread in the thread network (right column). All three discussion variables have a negative correlation with price volatility. The colors correspond to the level of the most closely related control variables (coin age, volume, and number of posts). For easier visualization, the control variables are represented as 4 equal-sized quantiles. We don't show plots from Design 2 since it is solely added as a robustness check against Design 1 and their plots look similar.

the announcement thread only a few days before trading, while a considerable number of altcoins start the discussion many weeks or even months prior to trading. As such, we limited our focus in Design 3 on altcoins which had a long enough discussion period to accurately reflect their collective sensemaking characteristics prior to trading.

6 RESULTS

6.1 Regression Analysis

Figure 5 shows the relationship between our three forum-based discussion variables and price volatility for Designs 1 and 3 as mentioned in Section 5. The scatter plots for both designs indicate that as contributors become more senior, the discussion becomes more diverse and its participation more uniform, the information available about the coin increases or similarly, the uncertainty around the coin (measured in terms of volatility) decreases. The only exception is the entropy of users' participation in Design 3 as it exhibits a positive relationship with volatility. Upon further analysis, we discovered this positive correlation in Design 3 is due to higher activity by the coin developers in the forum (a larger fraction of posts by a single user reduces the entropy). Discussion metrics in Design 3 are measured immediately after the coin is announced on the forum until the first trading date. During this period, high forum engagement by the announcers (development

Table 2. OLS Regression Results. All variables are scaled to zero mean and unit variance so that we can compare the magnitude of each coefficient. There are two models per each Design as described in section 5. The simpler model only includes the discussion variables, while the full model also includes the controls. The first two two columns correspond to Design 1 in which measurement period ends in November 2016. The second pair of columns correspond to Design 2 in which measurement period ends in January 2016. The third pair of columns represent Design 3 in which discussion (price) metrics are measured before (after) the coin starts trading. Age of the coin is measured in terms of the number of days the coin has been on the market until November 2016. Log of the total number of posts made by all users in the announcement thread acts a measure of total activity during the analysis period. The Bonferroni adjusted significance level for joint tests is 0.017 and 0.0084 in models with 3 and 6 variables respectively. At these significance levels, user age and entropy (almost) always reject the null, while the degree only rejects the null in models with 3 coefficients.

	Price Volatility over 100 Days					
	November 2016		January 2016		Initial Trading	
	(1)	(2)	(3)	(4)	(5)	(6)
Average User Age	-0.304^{***} p = 0.00005	-0.314^{***} p = 0.0001	-0.481^{***} p = 0.00000	-0.412^{***} p = 0.0002	-0.430^{***} p = 0.000	-0.442^{***} p = 0.0002
Number of Posts Entropy	-0.336^{***} p = 0.00001	-0.219^{**} p = 0.010	-0.236^{**} p = 0.007	-0.311^{***} p = 0.005	0.193^{***} p = 0.002	0.196^{**} p = 0.006
Thread Network Degree	-0.372^{***} p = 0.00000	-0.295^{*} p = 0.028	-0.211^{**} p = 0.007	0.060 p = 0.654	-0.226^{***} p = 0.0002	-0.296^{*} p = 0.023
Coin Age		0.062 p = 0.462		−0.162 p = 0.111		−0.081 p = 0.316
Average Daily Volume		-0.322^{***} p = 0.0003		−0.162 p = 0.084		0.051 p = 0.405
Number of Posts		0.092 p = 0.486		-0.302^{*} p = 0.049		0.080 p = 0.556
Observations	139	139	102	102	208	208
R^2	0.378	0.441	0.433	0.499	0.343	0.352
Adjusted R^2	0.364	0.416	0.416	0.467	0.333	0.333
Res. Std. Error	0.798	0.764	0.764	0.730	0.817	0.817
F Statistic	27.298^{***}	17.382^{***}	24.991^{***}	15.774^{***}	35.458^{***}	18.200^{***}
Note:						*p<0.05; **p<0.01; ***p<0.005

team) is crucial as it serves to inform the community about the coin's technical aspects. The developers' engagement in Design 3 is inherently more important than the first two designs in which the coin is already established. In fact, the correlation between the fraction of posts made by the developer and entropy is +0.64 ($p = 2.2 \times 10^{-16}$), and once we remove the posts made by the developer and recalculate the entropy, its relationship with volatility remains positive but weaker and barely significant.

Table 2 shows our main results from the regression analysis of price volatility according to two models for each of the three designs explained in Section 5.3. The first model in each design only includes the discussion variables while the second model also includes the controls as possible confounders. The entropy in Design 3 is computed including the posts made by the development

Table 3. The one-sided t-test for the difference of means between technical and non-technical coins for all the variables in design 1. The first row represents the observed empirical difference of means: $\overline{X}_{technical} - \overline{X}_{nontechnical}$. The second row shows the p-value of the Null hypothesis for the test H_0: *Non-Technical mean* > *Technical mean* for each variable except volatility. Rejecting the Null Hypothesis for our discussion metrics means that there is more collective sensemaking within the discussion of technical coins. In the case of volatility, the direction of the test is reversed *Non-Technical mean volatility* < *Technical mean volatility*. Rejecting the Null for volatility indicates that our operationalizations of information uncertainty are related as non-technical coins have higher volatility (uncertainty) than technical coins. Third row shows the 97.5% confidence interval for difference of means: $E[X_{technical}] - E[X_{nontechnical}]$

	Price Volatility	Log Daily Volume	Entropy of Posts	Log Thread Degree	User Ages	Coin Age	Log Total Posts
Means Diff.	-0.986	3.037	0.038	0.557	54.153	12.192	0.528
H_0 P-Value	0	6.09×10^{-5}	0.0185	0.00296	0.00704	0.453	0.102
97.5% CI	$(-\infty, -0.71)$	$(2.5, \infty)$	$(0.002, \infty)$	$(0.173, \infty)$	$(11.4, \infty)$	$(-194.4, \infty)$	$(-0.303, \infty)$

team in the announcement thread. We make several observations from the table. First, all discussion variable coefficients in all models, except for entropy in Design 3, are negative. Second, our discussion variables are remarkably significant even after controlling for all possible confounders in Design 1 (and Design 3) with only 139 (and 208) data points. Third, while the coefficient for the degree in Design 2 is significantly negative in the first model, it is no longer significant after including control variables. This is most likely due to the small sample size (only 102) and that the degree is highly correlated with the total number of posts in the thread. Finally, what matters is the seniority of the community around a coin, not the age of the coin itself: newly created coins can still attract senior and experienced users so long as they have less uncertainty and more information about their technical merits available.

As a robustness check and to validate these results, we performed the same analysis using a different operationalization for the level of information available about the coin, our technicality metric. Ideally, we should replicate our results using this binary indicator for technological innovation. However, since our sample size with this technicality indicator is too small, we cannot perform an exhaustive analysis such as Table 2. Instead, we perform multiple one-sided two-sample t-tests and MANOVA for the hypothesis that the technical coins with more objective information available exhibit larger entropy, degree, user age and less price volatility than non-technical coins.

Table 3 shows the result of these t-tests for all variables (i.e., volatility, discussion and control variables) using our data from Design 1. We make the following observations. First, non-technical coins have smaller volatility compared to technical coins indicating that the measure of innovation shares a significant commonality with our volatility measure. Second, the difference in means of all collective sensemaking variables is positive confirming the results in Table 2. Using the significance level of 0.025, we are able to reject the one-sided nulls for volatility and discussion variables. We have reported the p-values in table 3 so that Bonferroni corrections can be made easily depending on how many variables should be tested together, 3 or 6 by excluding or including the controls. The non-technical coins do not necessarily have a smaller level of discussion activity, pointing to the possibility of fads at least among a subset of these coins. A MANOVA test with a model that has entropy, degree and seniority as response variables and technicality as the grouping variable also indicates that the three discussion variables are significantly different between technical and non-technical coins ($p = 8.771 \times 10^{-5}$). Finally, Figure 7 illustrates the out-of-sample prediction power of the three discussion variables on 40% of Design 3 data as the test set.

The results above suggest there may be a subgroup of forum users who are pursuing new technical coins and actively participate in their discussions. Indeed, a cluster analysis of users based

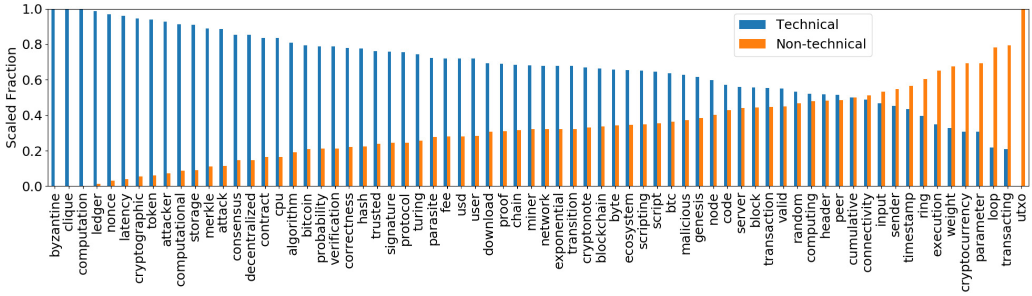


Fig. 6. The fraction of 70 most relevant words to cryptocurrencies vocabulary that appear in discussions of technical versus non-technical coins. For comparison, the fractions are scaled so that the technical and non-technical fractions sum up to 1 for each word sum. The distributions show that the relevant words more likely appear in technical coin discussions (Wilcoxon rank-sum test: $p = 1.0 \times 10^{-10}$). A detailed description of the analysis is explained in the Appendix A.3.

on their posting behavior confirms this finding. A k-means clustering of users based on how many posts they make in each announcement thread reveals that more than 85% of the all the posts in announcement threads of 10 most volatile coins are made by a single cluster of users. Similarly, more than 80% of the posts in announcement threads of 10 least volatile coins are made by two distinct clusters of users. In other words, the user bases of these two coin types are almost separate. Further supporting these results in Figure 6, we see that the technical coins, as described in section 4.3, are more likely to use the language specific to the cryptocurrency design than non-technical coins. The appendices A.2 and A.3 provides more detail on our clustering and text analyses[6].

6.2 Interviews

To qualitatively confirm and further investigate our findings, we conducted a set of structured interviews with participants on the forums. We designed several questions to obtain insights on how forum users perceive the differences in the discussion of technical versus non-technical altcoins. The Appendix A.4 lists these questions. We sent these questions to more than 100 senior users of the forum as direct messages, although we achieved a low response rate among those users, and we also published these questions as public threads on the forum. We were also able to directly communicate with the marketing head of a newly announced altcoin on the forum. In total, we received 5 replies including the single response from one of the veteran users via direct message.

Overall, the interviews provide qualitative verification of the importance of seniority, the only variable for which we have little theoretical justification due to limited previous work in the context of collective intelligence in social media. There is also some evidence on the importance of information diversity. Below, we list the main insights we got from the interviews along with essential quotes from the respondents:

(1) **User age** is a highly relevant variable indicative of expertise and mentioned by all respondents.

> *"Pump and dumps are here since the early days... you always have some folks buying a ton of XYZ s***coin - posting lots of "positive posts" with lots of "newbie" accounts in that coins thread and once people (mostly newbies) pick it up - they sell their coins."*
> *"Yes, newer members just say "wooooow, Ripple $100 next year", while experienced members tell them that's not possible."*

[6]The data and the code used to generate the results can be found at https://github.com/eamanj/cryptocurrency_discussions.

"Discussions that include a small number of active and experienced users are the best ones."
"The difference is certainly visible. In fact, the forum is not a lot of people, whose opinion is worth listening to. I think, those who are here for more than a year, have their own list of top 10-100 useful posts, topics, etc."
"I would think that the continued participation of particularly experienced users within the thread would be a strong marker of non-trivial coins."

(2) **Hype-based versus truth-seeking** discussions was a common concern of two respondents, including the marketing head of the newly launched altcoin. In their view, discussion of certain coins were highly substantial while a large number of discussions are superficial.

"The Ethereum announcement thread on Bitcointalk was full of relatively high quality discussion as to the merits or the inevitable failure of Ethereum as a project, but that was years ago. Now (especially after 2017) these threads are dominated by opportunists and speculators"
*"A lot of the ANN topics are filled with people managing several accounts and paid to bump the topic. Or most of the post are filled with s***posters posting "great project sir" "awesome project for the moon" etc... They are not posting because they're interested but because there is an incentive behind."*

(3) **Information diversity** was an important factor in sensemaking according to one of the respondents.

"I try to consider more than one source of information to make the analysis of the project more objective"

(4) **Continued engagement of the core users** in answering questions was mentioned by one respondent as an important factor in sensemaking of a project immediately after its launch. This is in line with our discussion of entropy in Design 3 mentioned in section 6.1.

"If in any topic a person begins to advertise the Scam project, I ask him about the reasons for choosing this project. As a rule, there is no answer."

6.3 Robustness Checks

Our work should be viewed as a descriptive analysis of the cryptocurrency community and the extent of self-deliberation or speculation in the community, and not a causal investigation. Our results indicate that when there is more information available about a crypto coin, the community discussion tends to also engage in more of a collective sensemaking role. In this context, our variables should be viewed as measures of truth-seeking-orientedness and measures of uncertainty about the technological value of a coin. Combined with the findings of [19], our results could suggest that there is at least a subset of users who are genuinely interested in discovering credible technological innovations.

Degree as a measure of diversity: In order to confirm that degree in the thread network indeed captures a notion of information diversity, we constructed two different, but more complex, measures of diversity among the users of each announcement thread. The first is the variation coefficient of participant ages within the coin announcement thread. The correlation between thread degree and this measure of age diversity is +0.31 ($p = 2.4 \times 10^{-4}$). The other diversity measure is based on the level of sentiment variation in posts of the announcement thread. For each post, we computed a sentiment score ranging from 0 (extremely negative) to 10 (extremely positive) using a dictionary-based sentiment scorer. The post sentiment score is the average sentiment score of the words

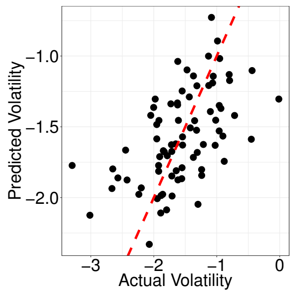


Fig. 7. Random forest prediction on test data from Design 3. 60% of data were used as training and the remaining 40% as test set. The correlation between actual versus predicted volatility is 0.545 with an R^2 of 0.296.

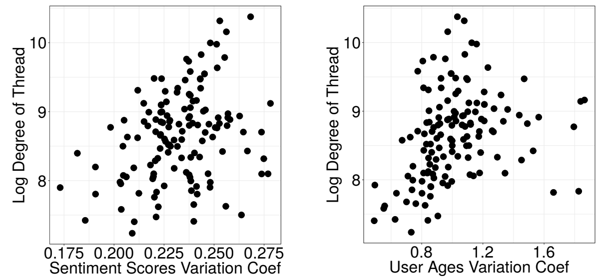


Fig. 8. The relationship between announcement thread degree and two alternative measures of diversity: sentiment and user age coefficient of variation in the announcement thread. The significant positive correlation (+0.32 on the left and +0.31 on the right) in both plots provides an extra evidence that thread degree measures another notion of (information) diversity.

that appear in the dictionary [42]. The coefficient of variation among all sentiment scores within an announcement thread serves as a measure of sentiment diversity. The correlation between thread degree and this measure of sentiment variation is +0.32 ($p = 1.5 \times 10^{-4}$). Figure 8 shows the relationship between thread degree and these two extra measures of discussion diversity.

Entropy of community participation: The results in section 6.1 indicate the sign of the correlation between the entropy of announcement posts and volatility depends on the measurement period. In particular, for coins that are already established (Designs 1 and 2), the correlation is negative, suggesting that less volatile coins exhibit discussion engagement by a broader user base and more equal community participation. On the other hand, the correlation is positive for newly launched coins (Design 3). We argued that this sign reversal happens because of active engagement by the development team to market their product and answer all concerns raised by the community. It appears to be important to secure active presence by a set of core users who promote it in the community. Figure 9 shows the relationship between the fraction of posts made by the announcer and the entropy of participation for Design 1 (established coins) and Design 3 (just launched). The announcer is often the developer or part of the development team. In Design 1, the posting activity by the coin announcer does not constitute a major part of the discussion (the fraction of posts made by the announcer is on average 6.2%). However, in Design 3, the announcer is an essential part of the discussion as they make on average 13.6% of the posts. The Spearman rank correlation between the fraction of posts made by the announcer and entropy is -0.24 ($p = 0.002$) in Design 1 and -0.52 ($p = 5.5 \times 10^{-15}$) in Design 3. These results suggest the developer engagement in the thread is a stronger driver of the behavior of entropy in Design 3 than Design 1.

To further investigate the opposite effects of the entropy metric, we removed the announcer and recalculated the entropy of announcement posts without the announcer. Figure 9 also shows the relationship between price volatility and this measure of entropy without the announcer. The strength of these relationships should be compared with that of Figure 5 in which entropy also includes the posts made by the announcer. The relationship in Design 1 is still strong and not affected by removing the announcer; however, it is much weaker than Figure 5 and barely significant in Design 3. There is still a positive correlation between this truncated measure of entropy and volatility in Design 3 because for many coins more than one user from the development team is

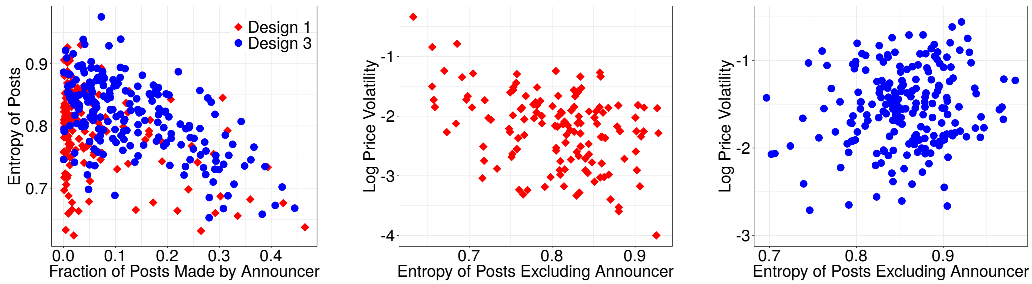


Fig. 9. The fraction of posts made by the announcer (first user in the announcement thread) versus normalized entropy of announcement posts (left). The announcers are more active prior to the launch. The entropy excluding the announcer versus volatility for Design 1 (middle) and Design 3 (right). The correlation in Design 1 is -0.40 ($p = 2 \times 10^{-6}$) comparable to -0.46 ($p = 9 \times 10^{-9}$) without removing the announcer. In contrast, the correlation in Design 3 is 0.15 ($p = 0.027$) significantly weaker than 0.30 ($p = 1.2 \times 10^{-5}$) without removing the announcer.

active in the announcement thread. For example, as mentioned in Section 5, there are at least 4 users from the XEM development team that are active on its announcement thread.

7 CONCLUSION

In the last year, cryptocurrencies have attracted massive attention from investors, institutions, policy-makers and the general audience. The public notoriety of Bitcoin, together with its sizable price increase [10], led to an explosion of attempts to create the *next Bitcoin*. Thus, a number of cryptocurrencies, often referred to as altcoins, and a vibrant set of exchanges have emerged particularly due to the extremely low cost and effort required to create or mutate a new coin, with some being minimal changes to parameters and branding of a pre-existing codebase. While many of these altcoins did not offer any new technological advancement, there have been some successful attempts in creating new cryptocurrencies that offered either significant technical innovation over the existing technology (e.g., Proof-of-stake in Peercoin) or introduced a wholly new idea (e.g. Turing Complete as in Ethereum) [43]. Given the abundance of new coins being created on a daily basis, it is natural to ask how well do traders detect cryptocurrencies that offer genuine technological innovation and are likely to succeed? A related question is whether the cryptocurrency community is attempting to collectively analyze and make sense of this large array of altcoins or is it simply engaged in hype-based speculation?

In this paper, we use an empirical approach to assess whether and when the discussions of cryptocurrencies are truth-seeking or hype-based. We rely on a novel data set that combines measures of the main online forum discussion around cryptocurrencies with their price and volume history in exchange markets. Leveraging the literature on finance, we assume price represents the perceived fundamental value of a coin and treat its volatility as an indicator of information uncertainty around the technological innovation of the cryptocurrency. Similarly, drawing upon collective intelligence literature and using three measures of experience, information diversity and (equal) community participation, we quantify the extent to which the community discussion exhibits characteristics of collective sensemaking.

Our results indicate a negative correlation between the quality of discussion measured in terms of collective sensemaking and price volatility of the coin suggesting that for "more serious" coins discussion is more likely to serve a truth-seeking role. Coins with more information available have equal participation by experienced contributors to the discussion and more diverse opinions measured in terms of access to other information sources. In contrast, coins with high information uncertainty tend to be discussed by less experienced and more narrowly focused users. We replicate the same results using an objective measure of technicality as a second operationalization of information uncertainty around the crypto coin. The content analysis of the forum also reveals that the discussion of more innovative coins is more focused on the design and technical aspects. These results are consistent with qualitative findings of [19] and suggest that there are people in the cryptocurrency community who are mainly driven by market hype and view cryptocurrency as an investment, while others are dedicated to the technological advancement of the cryptocurrency ecosystem and view Bitcoin and its variants as a legitimate currency.

Finally, we hypothesize that the same discussion patterns may also be present in other forms of social media. In order to distinguish between hype, fake news, and similar noise, one can look at the character of the discussion surrounding the news item, and in this manner filter out low-quality news items and promote those that exhibit characteristics of collective intelligence.

A APPENDIX

A.1 Supplementary Analysis

In this section, we discuss various robustness checks on our results by varying the data filtering criteria and the length of the time window over which variables were measured.

(1) **Data Filtering:** As mentioned in Section 5, a critical component of our analysis pipeline is filtering coins so that the coins in our analysis data have enough transaction volume and discussion activity during the study period in each Design. At low volumes, price data can be characterized as noise since a single small transaction can rapidly move the price. Similarly, low discussion activity does not hold any signals of collective information processing to be meaningful in our analysis. Therefore, we require the coins in Designs 1 and 2 (3) to have at least \$50 (\$150) of average daily volume over the 100 day analysis period. In addition, the coins also have to have at least 50 posts in their announcement thread during the forum analysis period in all Designs. A more relaxed filtering (lower thresholds) would introduce data points that likely resemble noise into the dataset, hence adversely affecting our results. On the other hand, a stricter filtering (higher thresholds) would limit the dataset to a smaller subset of more meaningful and higher quality data points. However, as the data size shrinks, our tests become less powerful. Nevertheless, it is important to evaluate the robustness of our results by checking whether similar results as in Table 2 hold if we loosen or tighten the filtering criteria. Tables 4 and 5 examine the robustness of our results by varying the filtering criteria in Designs 1 and 3 respectively. The results generally become weaker, nevertheless, they stay close to our findings in Table 2 with most p-values below 0.1 significance level in models with 6 variables. We believe the results confirm the robustness of our main finding in Table 2 to filtering criteria.

(2) **Time Window:** Another important parameter in our analysis is the length of price and discussion analysis periods (100 and 200 days respectively). In Section 5, we argued that the length of the discussion period should be long enough to accurately capture the health and substance of the community discussion patterns. Similarly, price analysis period should be long enough so we get an accurate estimate of average price volatility. Longer periods would increasingly incorporate shocks that are not the focus of our study and adversely affect our results. Nevertheless, we tested the robustness of our results by changing the length of these periods. We found that a 300 day period for discussion activity and a 50 day period for price volatility analysis lead to significant results similar to Table 2. However, discussion periods of 400 or 100 days led to much weaker results presumably due to mixing too many information shocks from the distant past or being too short to

Table 4. Robustness of our results to data filtering in Design 1. In loose (strict) filtering, average daily volume over 100 days should be greater than \$10 (\$250) and the announcement thread should have at least 5 (100) posts during the 200 days analysis period.

	Price Volatility over 100 Days			
	Loose Filtering		Strict Filtering	
	(1)	(2)	(3)	(4)
Average User Age	-0.306^{***}	-0.241^{***}	-0.396^{***}	-0.377^{***}
	p = 0.00000	p = 0.0001	p = 0.00001	p = 0.0002
Number of Posts Entropy	-0.291^{***}	-0.206^{***}	-0.270^{***}	-0.189
	p = 0.00001	p = 0.005	p = 0.002	p = 0.060
Thread Network Degree	-0.399^{***}	-0.161	-0.308^{***}	-0.354^{*}
	p = 0.000	p = 0.159	p = 0.0003	p = 0.035
Coin Age		-0.125		0.031
		p = 0.062		p = 0.762
Average Daily Volume		-0.355^{***}		-0.144
		p = 0.00000		p = 0.166
Number of Posts		-0.086		0.150
		p = 0.500		p = 0.348
Observations	220	220	106	106
R^2	0.288	0.416	0.353	0.374
Adjusted R^2	0.278	0.399	0.334	0.336
Residual Std. Error	0.850	0.775	0.816	0.815
F Statistic	29.113^{***}	25.251^{***}	18.576^{***}	9.866^{***}
Note:	*p<0.05; **p<0.01; ***p<0.005			

accurately capture the health status of community discussion. Similarly, a price analysis period of 200 days led to insignificant results, especially for Designs 1 and 2, potentially due to incorporating price shocks from the distant past.

A.2 Cluster Analysis

In this section, we describe the clustering analysis we briefly mentioned in Section 6. Our goal is to confirm whether there is a subgroup of forum users who are genuinely interested in discovering coins with new technological innovations. To this end, we use the subset of coins that appeared in Design 1 and cluster users based on the number of posts they made in each announcement thread. We first construct a matrix in which rows correspond to users (10,070 users) and columns correspond to announcement threads (139 coins). A cell (i, j) in the matrix denotes the number of posts user i has made in announcement thread of coin j during the 200 days prior to November 2016 (same as in Design 1). We then cluster the users using K-Means algorithm combined with Silhouettes scoring heuristic [48] to choose the optimal number of clusters. After clustering users

Table 5. Robustness of our results to data filtering in Design 3. In loose (strict) filtering, average daily volume over 100 days should be greater than \$60 (\$500) and the announcement thread should have at least 20 (100) posts during the 200 days analysis period.

	Price Volatility over 100 Days			
	Loose Filtering		Strict Filtering	
	(1)	(2)	(3)	(4)
Average User Age	-0.310^{***}	-0.380^{***}	-0.426^{***}	-0.465^{***}
	p = 0.000	p = 0.0001	p = 0.000	p = 0.001
Number of Posts Entropy	0.084	0.133^{*}	0.183^{**}	0.164^{*}
	p = 0.120	p = 0.039	p = 0.006	p = 0.035
Thread Network Degree	-0.190^{***}	-0.173	-0.264^{***}	-0.295^{*}
	p = 0.0005	p = 0.140	p = 0.0001	p = 0.036
Coin Age		-0.080		-0.074
		p = 0.260		p = 0.424
Average Daily Volume		-0.193^{***}		0.101
		p = 0.001		p = 0.135
Number of Posts		0.052		0.011
		p = 0.698		p = 0.939
Observations	311	311	164	164
R^2	0.154	0.189	0.361	0.374
Adjusted R^2	0.145	0.173	0.349	0.350
Residual Std. Error	0.924	0.910	0.807	0.806
F Statistic	18.581^{***}	11.793^{***}	30.178^{***}	15.656^{***}
Note:				*p<0.05; **p<0.01; ***p<0.005

into 7 distinct subgroups, we determine the fraction of posts made by users in each cluster for the 10 most and least volatile coins, where volatility is measured according to Design 1. Table 6 presents the fraction of posts made to these coins by the top 2 clusters contributing mostly to (non-)volatile coins. We conclude that there are distinct clusters of users who are exclusively active in either volatile or non-volatile coins (or potentially technical versus trivial coins). The results also suggest that there may exist significant assortativity [41] in the coins discussion network where edges correspond to the co-posting behavior of users in two announcement threads.

A.3 Content Analysis

In this section, we explain how the word distribution as depicted in Figure 6 was generated. To investigate how cryptocurrency communities were involved in a collective sensemaking process to learn about technical aspects of a coin, we did a simple content analysis of each coin's announcement thread. We analyzed the distributions of words relevant to cryptocurrencies vocabulary.

Table 6. The fraction of posts made in the announcement thread of 10 least (top) and most (bottom) volatile coins by their most contributing user clusters. The results suggest there are distinct user clusters that focus exclusively on either volatile or non-volatile coin discussions.

Coin Symbol	Price Volatility Over 100 Days	Fraction of Posts by Cluster 5	Fraction of Posts by Cluster 6
NVC	0.0254	0.9484	0.0023
DOGE	0.0280	0.0915	0.7488
LTC	0.0305	0.8187	0.0351
PPC	0.0339	0.9389	0.0075
DASH	0.0353	0.0359	0.0665
MONA	0.0408	0.0702	0.7703
XLM	0.0485	0.1514	0.0348
QRK	0.0511	0.6566	0.1415
XPM	0.0524	0.8804	0.0079
WDC	0.0531	0.8335	0.0833

Coin Symbol	Price Volatility Over 100 Days	Fraction of Posts by Cluster 0	Fraction of Posts by Cluster 1
YOC	0.3149	0.9775	0.0067
TRK	0.3325	0.7882	0.0926
SWING	0.3405	0.9858	0.0047
GLD	0.3428	0.0410	0.0062
1337	0.3639	0.9712	0.0000
EGC	0.3908	0.9554	0.0047
CRW	0.5512	0.9430	0.0129
SLING	0.6814	0.2441	0.7211
ARB	0.8685	0.9444	0.0336
8BIT	2.9277	0.8053	0.1340

We used the definition of technical (non-technical) coins as described in 4.3 and extracted relevant words from the announcement threads of these coins.

Relevant word extraction: We used a keyword extraction algorithm to define relevant keywords from online forum threads and to measure technicality score of each cryptocurrency discussion thread. For keyword extraction, we employed TF-IDF algorithm [50], which is commonly used in Information Retrieval and Natural Language Processing. TF-IDF algorithm defines a word score based on the multiplication of its term frequency (TF) in the document and its inverse document frequency (IDF) in the document collection. In our analysis, we calculated TF based on the term frequency of a word in a cryptocurrency corpus and IDF based on the document frequency of the word in a general corpus to detect keywords that appear exclusively in the discussion of cryptocurrencies. The TF-IDF score of a word w is calculated by

$$\text{TF-IDF}(w) = \log(\text{tf}_C(w) + 1) \cdot \log\left(\frac{N}{\text{df}_G(w)}\right), \tag{4}$$

where $\text{tf}_C(w)$ denotes the frequency of word w in a cryptocurrency corpus and $\text{df}_G(w)$ denotes the document frequency of word w in a general corpus. We used six technical papers of major cryptocurrencies (Bitcoin, Ethereum, IOTA, Ripple, Monero) as the cryptocurrency corpus to calculate $\mathrm{tf}_C(w)$. For $\mathrm{df}_G(w)$, we used Reuters and Brown[7] textual corpora that contain general vocabularies not related to cryptocurrency discussions. After extracting relevant keywords using TF-IDF scoring, we further trimmed them manually to ensure they have a technical meaning pertaining to design aspects or business of cryptocurrencies.

Word distribution calculation: For each group of coins (i.e., technical coins and non-technical coins), we extracted all posts that were published in the announcement threads of the coins from 2016-04-15 to 2016-11-01. We filtered out the posts by the first user of each thread who announced the coin to remove the bias in vocabulary usage of the cryptocurrency authors. We also removed any coins whose discussion thread contains less than 50 posts to avoid involving inactive discussions. The remaining posts of each coin group were concatenated into a single document. For each coin group, we calculated the fraction of each word (thus, the fractions of all words sum up to 1). All words were tokenized and lemmatized (e.g., Both "nodes" and "node" were canonicalized into node) using RegexTokenizer and WordNetLemmatizer of NLTK Library [8]. Then, we compared the distributions of the relevant keywords defined above between technical coin discussions and non-technical coin discussions. This comparison is shown in Figure 6 which suggests that the discussion of technical coins with more public information available is more likely to contain the vocabulary exclusive to the cryptocurrency design.

A.4 Interview Questions

In this section, we list the exact questions we asked forum users either in a public thread or through direct messages.

(1) If you were to predict the eventual success of a new altcoin, would the discussion patterns in its announcement page provide any meaningful signal? If yes, how so?
(2) What are good marketing strategies on bitcointalk?
(3) Have you noticed any difference in discussion between members who have been on bitcointalk for many years versus just a few weeks? Or communities that focus on technological innovation versus those who only care about identifying a good investment?
(4) Is there any specific patterns in discussions that are likely to be related to fraudulent coins or pump and dump schemes?
(5) Finally, I am interested if you have any insights on the difference between the ANN page of highly technical coins (like ETH or Monero) vs trivial coins that just changed a few parameters. This could be anything, some examples are content of discussions, experience of users, length of discussions or the social relationship of users in the discussion.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. This works is also supported in part by the MIT Trust Data Consortium, the MIT Media Lab Members Consortium and Ministerio de Economía y Competividad (Spain) through projects FIS2013-47532-C3-3-P and FIS2016-78904-C3-3-P. We would like to give special thanks to Nicolás Della Penna for helping to supervise the initial data collection.

[7] http://www.nltk.org/nltk_data/
[8] https://www.nltk.org/

REFERENCES

[1] Jonas Valbjørn Andersen and Claire Ingram Bogusz. 2017. Patterns of Self-Organising in the Bitcoin Online Community: Code Forking as Organising in Digital Infrastructure. *ICIS* (2017).

[2] Denise Anthony, Sean W Smith, and Tim Williamson. 2005. Explaining quality in internet collective goods: Zealots and good samaritans in the case of wikipedia. *Hanover: Dartmouth College* (2005).

[3] Geert Bekaert, Eric Engstrom, and Yuhang Xing. 2009. Risk, uncertainty, and asset prices. *Journal of Financial Economics* 91, 1 (2009), 59 – 82. http://www.sciencedirect.com/science/article/pii/S0304405X08001670

[4] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2, 1 (2011), 1–8.

[5] Eric Bonabeau and Christopher Meyer. 2001. Swarm intelligence: A whole new way to think about business. *Harvard business review* 79, 5 (2001), 106–115.

[6] Jamal Bouoiyour and Refk Selmi. 2015. What does Bitcoin look like? *Annals of Economics & Finance* 16, 2 (2015).

[7] Eng-Tuck Cheah and John Fry. 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters* 130 (2015), 32–36.

[8] Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27, 5 (2014), 1367–1403.

[9] Jonathan Chiu and Thorsten V Koeppl. 2017. The economics of cryptocurrencies–bitcoin and beyond. (2017).

[10] coinmarketcap. [n. d.]. coinmarketcap.com. https://coinmarketcap.com

[11] Michael A Cusumano. 2014. The bitcoin ecosystem. *Commun. ACM* 57, 10 (2014), 22–24.

[12] David M Cutler, James M Poterba, and Lawrence H Summers. 1988. What moves stock prices?

[13] Alexander David and Pietro Veronesi. 2000. Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities. 47 (2000).

[14] Nathan Eagle, Michael Macy, and Rob Claxton. 2010. Network Diversity and Economic Development. *Science* 328, 5981 (2010), 1029–1031. https://doi.org/10.1126/science.1186605

[15] Abeer ElBahrawy, Laura Alessandretti, Anne Kandler, Romualdo Pastor-Satorras, and Andrea Baronchelli. 2017. Evolutionary dynamics of the cryptocurrency market. *Royal Society Open Science* 4, 11 (2017). https://doi.org/10.1098/rsos.170623 arXiv:http://rsos.royalsocietypublishing.org/content/4/11/170623.full.pdf

[16] David Engel, Anita Williams Woolley, Ishani Aggarwal, Christopher F. Chabris, Masamichi Takahashi, Keiichi Nemoto, Carolin Kaiser, Young Ji Kim, and Thomas W. Malone. 2015. Collective Intelligence in Computer-Mediated Collaboration Emerges in Different Contexts and Cultures. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3769–3778. https://doi.org/10.1145/2702123.2702259

[17] David Engel, Anita Williams Woolley, Lisa X Jing, Christopher F Chabris, and Thomas W Malone. 2014. Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PloS one* 9, 12 (2014), e115212.

[18] Neil Gandal, JT Hamrick, Tyler Moore, and Tali Oberman. 2017. Price Manipulation in the Bitcoin Ecosystem. *Journal of Monetary Economics* (2017).

[19] Xianyi Gao, Gradeigh D. Clark, and Janne Lindqvist. 2016. Of Two Minds, Multiple Addresses, and One Ledger: Characterizing Opinions, Knowledge, and Perceptions of Bitcoin Across Users and Non-Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1656–1668. https://doi.org/10.1145/2858036.2858049

[20] David Garcia and Frank Schweitzer. 2015. Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science* 2, 9 (2015). https://doi.org/10.1098/rsos.150288

[21] Eric Gilbert and Karrie Karahalios. 2010. Widespread Worry and the Stock Market. In *ICWSM*. 59–65.

[22] Martine Haas and Mark Mortensen. 2016. The secrets of great teamwork. *Harvard business review* 94, 6 (2016), 70–6.

[23] Adam S Hayes. 2017. Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics* 34, 7 (2017), 1308–1321.

[24] Lu Hong and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America* 101, 46 (2004), 16385–16389. https://doi.org/10.1073/pnas.0403723101

[25] Lin Hsiu-Fen and Lee Gwo-Guang. 2006. Determinants of success for online communities: an empirical study. *Behaviour & Information Technology* 25, 6 (2006), 479–488. http://libproxy.mit.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=23148703&site=ehost-live&scope=site

[26] Investopedia. 2018. Intrinsic Value. https://www.investopedia.com/terms/i/intrinsicvalue.asp

[27] Jermain Kaminski. 2014. Nowcasting the bitcoin market with twitter signals. *arXiv preprint arXiv:1406.7577* (2014).

[28] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.

[29] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: Enhancing group collaborationusing sociometric feedback. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. ACM, 457–466.

[30] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one* 11, 8 (2016), e0161197.

[31] Young Bin Kim, Jurim Lee, Nuri Park, Jaegul Choo, Jong-Hyun Kim, and Chang Hun Kim. 2017. When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation. *PloS one* 12, 5 (2017), e0177630.

[32] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W Malone. 2017. What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends.. In *CSCW*. 2316–2329.

[33] Aniket Kittur, Ed Chi, Bryan A Pendleton, Bongwon Suh, and Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web* 1, 2 (2007), 19.

[34] Aniket Kittur and Robert E. Kraut. 2010. Beyond Wikipedia: Coordination and Conflict in Online Production Groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 215–224. https://doi.org/10.1145/1718918.1718959

[35] Peter M Krafft, Nicolás Della Penna, and Alex Pentland. 2018. An Experimental Study of Cryptocurrency Market Dynamics. *ACM CHI Conference on Human Factors in Computing Systems (CHI)* (2018).

[36] Stefan Krause, Richard James, Jolyon J Faria, Graeme D Ruxton, and Jens Krause. 2011. Swarm intelligence in humans: diversity can trump ability. *Animal Behaviour* 81, 5 (2011), 941–948.

[37] Ladislav Kristoufek. 2013. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports* 3 (2013), 3415.

[38] M. Linton, E. G. S. Teo, E. Bommes, C. Y. Chen, and Wolfgang Karl Härdle. 2017. *Dynamic Topic Modelling for Cryptocurrency Community Forums*. Springer Berlin Heidelberg, Berlin, Heidelberg, 355–372.

[39] Martina Matta, Ilaria Lunesu, and Michele Marchesi. 2015. Bitcoin Spread Prediction Using Social and Web Search Media.. In *UMAP Workshops*.

[40] Audris Mockus, Roy T. Fielding, and James D. Herbsleb. 2002. Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Trans. Softw. Eng. Methodol.* 11, 3 (July 2002), 309–346. https://doi.org/10.1145/567793.567795

[41] M. E. J. Newman. 2002. Assortative Mixing in Networks. *Physical Review Letters* 89, 20 (Oct. 2002), 208701. https://doi.org/10.1103/PhysRevLett.89.208701

[42] F. Å. Nielsen. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on Making Sense of Microposts (MSM)*. 93–98. http://arxiv.org/abs/1103.2903

[43] Bobby Ong, Teik Ming Lee, Guo Li, and David LEE Kuo Chuen. 2015. Evaluating the potential of alternative cryptocurrencies. In *Handbook of digital currency*. Elsevier, 81–135.

[44] F. Ortega, J. M. Gonzalez-Barahona, and G. Robles. 2008. On the Inequality of Contributions to Wikipedia. (2008), 304–304.

[45] Jose Pagliery. 2014. *Bitcoin: And the Future of Money*. Triumph Books.

[46] Michal Polasik, Anna Iwona Piotrowska, Tomasz Piotr Wisniewski, Radoslaw Kotkowski, and Geoffrey Lightfoot. 2015. Price fluctuations and the use of Bitcoin: An empirical inquiry. *International Journal of Electronic Commerce* 20, 1 (2015), 9–49.

[47] O. Poyser. 2017. Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series. *ArXiv e-prints* (June 2017). arXiv:1706.01437

[48] Peter Rousseeuw. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* 20, 1 (Nov. 1987), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

[49] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854–856.

[50] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. 2008. *Introduction to information retrieval*. Cambridge University Press.

[51] Lynette Shaw. 2016. *The Meanings of New Money: Social Constructions of Value in the Rise of Digital Currencies*. Ph.D. Dissertation.

[52] Yhlas Sovbetov. 2018. Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero. 2 (02 2018), 1–27.

[53] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi. 2010. Want to Be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing (SOCIALCOM '10)*. IEEE Computer Society, Washington, DC, USA, 177–184. https://doi.org/10.1109/SocialCom.2010.33

[54] Paul C Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* 62, 3 (2007), 1139–1168.

[55] F. Tschorsch and B. Scheuermann. 2016. Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies. *IEEE Communications Surveys Tutorials* 18, 3 (thirdquarter 2016), 2084–2123. https://doi.org/10.1109/COMST.2016.2535718

[56] Andrew Urquhart. 2016. The inefficiency of Bitcoin. *Economics Letters* 148 (2016), 80–82.

[57] Paul Vigna and Michael J Casey. 2016. *The age of cryptocurrency: how bitcoin and the blockchain are challenging the global economic order.* Macmillan.

[58] Anita Williams Woolley, Ishani Aggarwal, and Thomas W Malone. 2015. Collective intelligence and group performance. *Current Directions in Psychological Science* 24, 6 (2015), 420–424.

[59] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (2010), 686–688. https://doi.org/10.1126/science.1193147 arXiv:http://science.sciencemag.org/content/330/6004/686.full.pdf

[60] Aaron Yelowitz and Matthew Wilson. 2015. Characteristics of Bitcoin users: an analysis of Google search data. *Applied Economics Letters* 22, 13 (2015), 1030–1036.

Received April 2018; revised July 2018; accepted September 2018