Forever Blowing Bitcoins: Social Structure and Speculative Bubbles in Cryptocurrencies

Nicolas Della Penna [*]
ANU
n@nikete.com

Eaman Jahani [†], Peter Krafft [‡]
MIT
eaman,pkrafft@mit.edu

Octavio Bunge [§]
Universidad de Belgrano
octavio.bunge@comunidad.ub.edu.ar

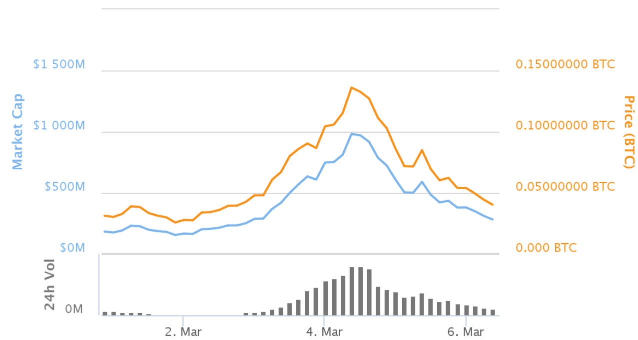
Harper Reed [¶]
harper@nata2.org

Alex (Sandy) Pentland
MIT Media Lab
sandy@media.mit.edu

Julian McAuley
UC San Diego
jmcauley@cse.ucsd.edu

ABSTRACT

We study the power of structural features of the social network around cryptocurrencies to understand the severity and magnitude of bubbles. Novel dataset contribution summary line. TODO. All our measures are constructed on the state social network before the relevant cryptocurrency is ever traded.



1. INTRODUCTION

Speculative bubbles are perceived to periodically take over markets [**?**]. Going back at least to the *South Sea* bubble in the early 18th century, well-informed parties have invested knowingly in bubbles, and found it profitable [**?**]. Today, the public notoriety of Bitcoin, together with its massive price increases and their associated publicity lead to an explosion of attempts to create "the next Bitcoin". Collectively, these currencies are often referred to as "cryptocurrencies" or simply "coins", and a vibrant set of exchanges have emerged where these are traded, either for each other or money. The majority of these coins have no possible future valuable in the long term, and their markets would appear to be driven largely by speculation. Many of them appear to be nothing but attempts at turning a quick profit from inflating the implied valuation of a coin shortly after creating it. This is driven by the extremely low cost and effort required to create a new coin, with most being minimal changes to parameters and branding of a pre-existing codebase.

Those who make and trade these coins communicate largely online, and much of their activity is concentrated on public forums. As such, price and volume data from these exchanges is freely available and widely aggregated[1], and the source code to all of them being public is the normThis makes cryptocoins an ideal lens through which to study the social life of a "market mania" [**?**]. Such a study is valuable both as a means of understanding the dynamics of bubbles themselves, but also from a computational social science perspective, to understand the ecosystem and lifecycles of these online communities.

We present a study of a large number of cryptocurrency ecosystems, using a novel dataset that combines measures derived from social networks of users in cryptocurrency forums, market data aggregated over dozens of exchanges, and properties of the software implementing the cryptocurrencies themselves. From forums we identify the introducers of each coin and build measures of their position in the network based on their patterns of engagement in forum threads *before* the coin is announced. In this way we identify 376 coins that are announced by users of the forum and which can be mapped to price and volume data from exchanges. Next, from price and transaction volume data we build measures of the subsequent activity that results from trading in the coin. We also assess if coins possibly embody technological innovation based on having more than trivial modifications to previously existing coins' source code.

[1] these are however largely unregulated, often anonymous, and there is no way to account how much of reported volumes are manipulation attempts. The existence of attempts at arbitrage between them places some bounds on how blatantly the data can be manipulated for prices, but volumes are impossible to assess.

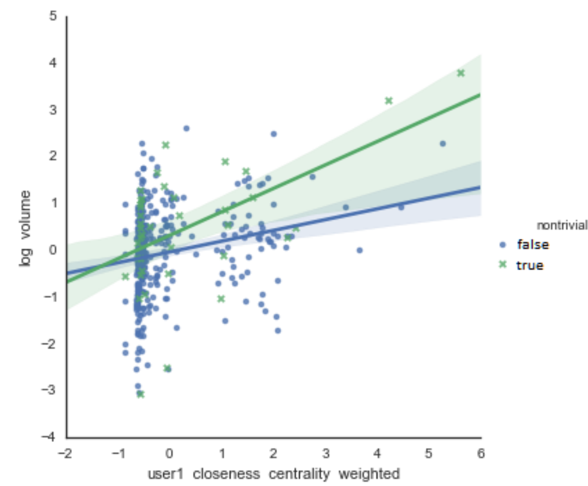
While the mechanisms that drive bubbles have been studied both theoretically [?, ?, ?, ?] and experimentally [?], an exhaustive dataset and study of the social networks of those promoting the asset has not been previously previously possible. While the magnitude of the assets traded is certainly small relative to most financial and commodity markets, it is nevertheless much larger than even the most lavishly funded experimenter could hope for. For example, the largest bubble in our dataset, AuroraCoin, reaches a valuation of one billion USD in March 2014, with reported daily trading volumes of 6.8M USD, and sheds 90% of its value in a week, and 99% of its value in well under a year. For context, this is equivalent to one quarter of Iceland's entire foreign exchange reserves in 2014,[2] the population of which AuroraCoin promoters claimed they would distribute half of the coins to.

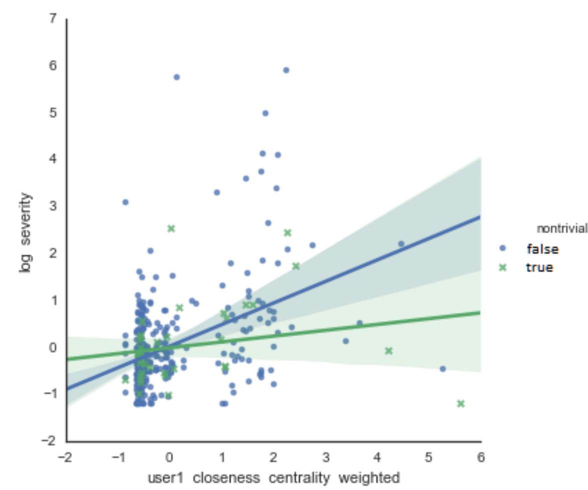
Using price and volume data we construct measures of both the magnitude and the severity of bubbles. These are defined formally in our variable section, but their intuition is that we say the magnitude of a coin is large when a high volume in dollars of trades happens, while we say a coin has a severe bubble if investing a fixed amount leads to loosing a large proportion. By considering the community structure that exists in the forum before a coin is introduced we are able to predict a substantial fraction of the variation in both the severity and magnitude of the resulting bubble. This is a challenging task, and models that rely on either simple activity or network metrics metrics show almost no predictive out-of-sample power, unable to explain even 1% of the variation in either task, our best models perform an order of magnitude better in both tasks. The main driver of our explanatory power is the centrality of a user in the directed network derived from the forum. This effect appears to be mediated by whether a coin involves a nontrivial technological change, the direction of the interaction reversing depending on whether it relates to magnitude or severity. Both the severity and the magnitude of bubbles increases with the centrality of the user who introduces the coins in the forums. Interestingly this effect is concentrated in different ways depending on whether the coin software is more than a trivial modification: trivial coins have more severe bubbles the more central their introducers are, while volume is greater the more central the introducer of a nontrivial coin is.

2. LITERATURE

This work is at the intersection of three literatures: in economics and finance on the study of speculative bubbles, in network science on the prediction of outcomes based on features of an individual in a network, and in computer science, largely centered on the security community, studying cryptocurrencies.

While states can create the demand required for a currency system to run by compelling tax payment in it (for a recent example), non-state-sponsored currencies must find some other ways of creating demand. The initial market for which bitcoin has been used (prices denominated in it, transactions only in it) was drug sales.[3] Since the cost of producing a





[2]4.1 Billon USD, The World Bank, Global Economic Monitor, accessed October 2015

[3]A overview of the different drug marketplaces and esti-

new coin is effectively zero, new currencies have thus been floated with every single drug name possible. Many chains can claim to the same name, so exchanges with volume (since speculation is the only possible use of almost all of the coins) become de-facto arbitrators of who has a minimally viable claim.[4]

2.1 Bubbles and Herds

Perhaps the most striking line of research on bubbles in economics with respect to cryptocurrencies is the study of markets where the asset is worthless and this is common knowledge. Recently [?] studies both theoretically and experimentally in the laboratory such a bubble. The driving force is that some traders "do not know where they stand in the market sequence, the game allows for a bubble at the Nash equilibrium when there is no cap on the maximum price.". In the context of cryptocurrencies the lack of knowledge around the sequence position maps to uncertainty about ones place in the technology adoption knowledge and adoption curve, while the difficulty in upper bounding the potential market value of cryptocurrencies provides the lack of cap on the maximum price.

A large literature in finance empirically examines herding by financial analysts, for a recent example[?] tests the hypothesis of herding in analysts forecasts.

It also looks at the properties of analysts who disagree with their peers and of their forecasts

[?] weather analysts are likely to disagree with their peers for a prominent recent example classifies analysts' earnings forecasts as herding or bold and finds that (1) boldness likelihood increases with the analyst's prior accuracy, brokerage size, and experience and declines with the number of industries the analyst follows, consistent with theory linking boldness with career concerns and ability; (2) bold forecasts are more accurate than herding forecasts; and (3) herding forecast revisions are more strongly associated with analysts' earnings forecast errors (actual earningsâĂŤforecast) than are bold forecast revisions. Thus, bold forecasts incorporate analysts' private information more completely and provide more relevant information to investors than herding forecasts."

2.2 Bitcoin and Cryptocurrencies

[?, ?]

heuristic clustering to group Bitcoin wallets based on evidence of shared authority, and then using re-identification attacks (i.e., empirical purchasing of goods and services) to classify the operators of those clusters. From this analysis, we characterize longitudinal changes in the Bitcoin market, the stresses these changes are placing on the system, and the challenges for those seeking to use Bitcoin for criminal or fraudulent purposes at scale."

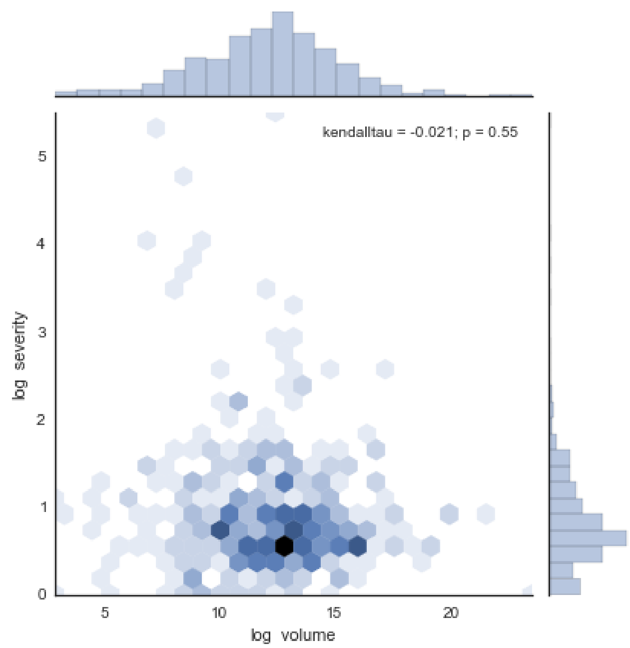
[?] fistful of bitcoins Bitcoin is a purely online virtual currency, unbacked by either physical commodities or sovereign obligation; instead, it relies on a combination of cryptographic protection and a peer-to-peer protocol for witnessing settlements. Consequently, Bitcoin has the unintuitive property that while the ownership of money is implicitly anonymous, its flow is globally visible. In this paper we explore this unique characteristic further, using heuristic clustering to group Bitcoin wallets based on evidence of shared authority, and then using re-identification attacks (i.e., empirical purchasing of goods and services) to classify the operators of those clusters. From this analysis, we characterize longitudinal changes in the Bitcoin market, the stresses these changes are placing on the system, and the challenges for those seeking to use Bitcoin for criminal or fraudulent purposes at scale.

[?] Measuring the longitudinal evolution of the online anonymous marketplace ecosystem

[?] How Did Dread Pirate Roberts Acquire and Protect His Bitcoin Wealth

3. DATA DESCRIPTION

3.1 Prices, Exchanges, and Coin characteristics



Our main outcome measures are the severity of drop in the value of a unit of the asset, and the magnitude in USD of the transactions in it. We scrape daily price and volume data from coinmarketcap.com[5] We operationalize the severity of a bubble as the inverse of 1 dollar that would be lost buying at the maximum price and selling after that proportionally

mated transaction volumes can be found in [?]. To the best of the authors' knowledge no other sector beyond speculation has even remotely substantial volume at present; a very primitive form of unregulated gambling Satoshi Dice, did for a brief pointing the past)

[4]While it is theoretically possible to engage in a distributed protocol to exchange between two cryptocurrencies, see part II of lecture 10 in [?]

[5]For robustness analysis smaller subsets of the coins where available from coin

to the volume of the market until the present; we call this severity. We define the volume as the sum of the dollar volume of trade reported over all exchanges.

3.2 Forum Discussions

In order to study the effect of communication networks around cryptocoins on price variations, we collected all the posts from the most popular cryptocurrency online community, bitcointalk. Our data consists of all the posts that were made between January 2010 and July 2015 on the most active crypto-related forums:

1. **Bitcoin Discussion:** This is the oldest forum on the website which mainly focuses on issues related only to Bitcoin. Interestingly, Satoshi Nakamoto, the alleged creator of Bitcoin made the first post on this forum in January 2010 and was active until January 2011. The presence of Satoshi in the data set enables us to study the position of various actors in the online community relative to Satoshi and its relation with the success or failure of cryptocoins they advocate or reject.
2. **Altcoin Discussion:** This is the most active forum in the community with more than 730,000 posts as of July 2015, dating back to June 2011. The discussions in this forum mainly evolve around alternative currencies other than Bitcoin. Users often discuss the merits or flaws of various altcoins or simply exchange technical information.
3. **Announcement (Altcoin):** Community announcements such as development of exchange clients or addition of new features are made here. This is an important forum in our study as the creation of new altcoins are announced here. Whenever a new altcoin is announced to the community, the announcement is tagged with string ANN. This enables us to detect announcements of new coins into the market and identify the users who introduced them for the first time.
4. **Mining (Altcoin):** Technical issues pertaining to mining (i.e. validating transactions) altcoins are discussed here.
5. **Marketplace (Altcoin):** This forum contains the discussions on a wide-range of market-related issues, such as price or volume trends, possible pump-and-dump schemes and exchange tips.

Each forum consists of many subjects or threads initiated by different users. Each thread contains several posts or replies, with an average of 10 posts per thread. The reply structure within each thread constitutes the basis of our forum network, discussed below. Each post has several fields which contain valuable information in our context.

1. **Subject:** Usually, the initiator of the thread chooses subject and all the following posts inherit the same subject.
2. **Content:** The actual text of the post.
3. **Position in the thread**: The later posts in the thread might not be as important as earlier posts and could be about issues other than the original topic of the thread.
4. **Author**
5. **Date**

The community had only 10,000 unique users until early 2013, however it grew considerably faster after 2013 and reached about 70,000 by early 2015. Ultimate there are around 10,000 active users within a 30 day period on average.

4. FORUM INTERACTION NETWORK

Given the forum discussion data at the level of individual posts, we construct a network capturing the discussion patterns among users. The structural properties of this network form the basis of our analysis on a per-coin basis. In this network, nodes are the forum users and *directed edges* point from posters within each thread to thread-initiators. The omission of edges based on simple co-appearance within a thread leads to a sparser network which isolates the communication patterns around "dialogue-shapers". The edges in the discussion network are weighted by the number of times a poster replies to a thread-initiator in different threads (i.e. multiple replies by the same user within the same thread are counted only once). In this context, edge weights capture the level of engagement thread initiators receive from the community and the amount of information a poster receives from thread initiators. Furthermore, our network construction method uses all the interactions since the inception of bitcointalk in creating new edges or updating their weights. The unlimited retention of any such (replier to thread initiator) interaction captures relevant information on seniority and community influence which are obtained through long-term and persistent presence in the forums.

To build our network, we first combine the posts fro all forums. We do this because the community base of all five forums mentioned above is made of the same users, and because we are mostly concerned about influence and aggregate information flow among users, rather than the exact topic of the discussion. The network construction involves replaying all the posts over time sorted by their date and updating the discussion network accordingly. Whenever a new altcoin is introduced in the forum for the first time, the user who introduced it and a snapshot of the network is taken. We analyze the discussion network only up to the first time each coin is introduced to the community, in order to avoid any possible confounding between a coin's price movement and the extra attention it receives in the community due the same price changes. Our method uses the position of the first introducer in the network snapshot and the general structure of her neighborhood for extracting various network measures corresponding to that coin. Our final analysis examines these per-coin measures for evaluating the performance of each coin.

The majority of such introductions are made in the *Announcement* forum and are preceded with the "ANN" tag. We look for the first mention of both the coin symbol **and** its descriptive name in the subject of a thread which contains the announcement tag. The first mentions of either the coin symbol **or** its name are used as a fall-back in case the more restrictive **and** requirement did not detect the coin. Using this method, we were able to detect the first introduction of 554 altcoins out of 679. The forum user who initiated such a thread is assigned as the introducer of the coin to the community.

TODO(NIKETE): add validation results table wrt mapofcoins data here

4.1 Nontrivial coins

Many of the coins available in the exchanges are trivial modifications of another coin in that they only change parameters such as the name, the number of total mineable coins, or the transaction time between blocks. These coins' production cost is virtually zero[6].

To attempt to capture this we analyze data from mapofcoins.com which includes a genealogy of coins and data from the github page of coins not available on mapofcoins.com. If the coin to be analyzed has a parent and the algorithm it uses differs from the parent or if it has no parent, it is labeled as nontrivial, meaning that the coin implemented something that did not previously exist, and is not just a fork with only parameters changed, such as total mineable coins, transaction speed, etc.

4.2 Network Structure

In this section, we discuss the various metrics extracted from discussion networks and used as independent variables in the regression analysis. Many of these variables are standard metrics in graph theory designed to capture node centrality is specific scenarios [?]. As mentioned before, each coin is associated with a forum user and a discussion network which corresponds to the state of the forum at the time the user introduced the coin to the community. All of our node-level variables refer to the user introducing the coin. Below, we list the network variables included in the analysis. We used Python igraph implementation for computing the network-related metrics [?].

1. **Introducer number of posts:** The total number of posts (thread-initiations or simple replies) the coin introducer has made at the time she introduces the coin. It captures the user's level of activity in the community.
2. **Introducer number of threads:** The total number of threads the coin introducer has made. Users who start many threads are more likely to receive incoming edges and to shape the dialogue in the community.
3. **Seniority:** It is the number of days since the user's first post in the forums. We use this as a proxy for user's seniority in the community.
4. **Incoming degree:** The (incoming) degree centrality captures the role of dialogue-shapers in the community as it is the number of unique users who have replied to any of the focal user's threads.
5. **Outgoing degree:** The (outgoing) degree centrality captures the role followers in the community as it is the number of unique thread initiators the focal user has ever replied to.
6. **Total degree:** The (undirected) degree centrality captures the total level of the user's involvement in the community in any of the two forms above.
7. **Clustering Coefficient:** A measure of embeddedness or triadic closure, this is the fraction of the focal user's triads that are closed. In general, ideas are more likely to be reinforced and persistent in a triad if it is 'tightly-knit'. The positive effect of triadic closure (and balance) on tie qualities and their persistence is shown to exist in online social network such as Twitter [?], and we believe the same argument applies to our scenario.
8. **Unweighted closeness centrality:** While degree centrality measures the level of user engagement in the community, it only examines the local structure around the user. In contrast, closeness centrality measures the level of the user's engagement with the global network *either directly or indirectly.* It is relevant in many scenarios, including in online discussions, as information spreads via shortest paths.

   In our context, a user with high closeness centrality has interacted with a diverse set of users who themselves are close to a large set of diverse users. Our analysis consisted of three versions of the unweighted closeness centrality:

   (a) **Incoming:** Only the directed paths leading to the focal user are used. It measures closeness of the whole network to the focal user. Users who start many threads are likely to have higher incoming closeness centrality.

   (b) **Outgoing:** Only the directed paths starting from the focal user to all the other users are used. It measures closeness of the focal user to the whole network. Users who reply to many threads are likely to have higher outgoing closeness centrality.

   (c) **Undirected:** The paths both from and to the focal users are used. Users who initiate and reply to many threads are likely to have higher undirected closeness centrality.

   All measures are normalized by the number of users present in the network at the time of coin introduction, so that the comparison between the closeness centrality of various users (who introduce the coins) at different times is valid.
9. **Weighted closeness centrality:** Similar to the unweighted closeness centrality above, we computed three different versions, with the exception that edges are weighted to indicate the intensity or level of interaction between the two users. The edge weights in our discussion network are determined by the frequency of interactions between two users; and as two users interact more, they are deemed to be closer in their shortest path. Thus in the computation of weighted closeness centralities, we use the reciprocal of the weights as the distance between two users.

$$C_i = \frac{N-1}{\sum_{j=1}^{N} \sum_{e \in S_{ij}} \frac{1}{w_e}} \qquad (1)$$

where e denotes an edge in S_{ij} the set of users on the shortest path from i to j. w_e is the weight of edge e determined by the number of interactions between the end points.

[6] there is a cottage industry that offers the creation of binaries and provisioning of mining and bandwidth as a bundled service that require no technical skill to create form the user, examples are Coingen or Coincreator

10. **Weighted betweenness centrality:** Betweenness is closely related to the theory of weak ties and structural holes and measures how well of a bridge is the focal user. In our context, one could interpret betweenness centrality as a generational bridge. Bitcointalk has been an active forum since early 2010, and many users who were once active in its early days are no longer present in the forum. There are however some early users who are still active on the forum. These users have high betweenness centrality as they act as generational bridges between founders and the newcomers to the community. Another standard interpretation of high betweenness centrality is the existence of users who simulatenously interact with two isolated communities in the forum. Similar to closeness centrality, our betweenness centrality computation uses the inverse of edge weights as the distance between two users.
11. **Satoshi distance:** Distance to Satoshi can be interpreted as founder effect. Closeness to or direct interaction with Satoshi constitutes as a form of social capital in the community.
12. **Weighted pagerank:** where higher weights measured by the frequency of interactions facilitate the flow.
13. **Weighted Satoshi pagerank:** is similar to regular pagerank above with the only difference that resets of the random walk always direct to Satoshi instead of a uniform distribution over all users. It can be interpreted as the level of influence or creditability allocated from the founder to other users.

5. METHODS

Initially we we start with a baseline model that considers only the user characteristics that are easily observable from their activity on the forum before the announcement: the number of posts and of subjects, the time since they first post, the number of users that they have responded to and received responses from. These network measures are possible for any generic discussion, we introduce two further sets of variables to enrich our models that rely on domain knowledge of the underlying assets: Satoshi network measures, and weather a given coin is embodied in new software or if it is simply a change in name and parameters of the codebase used by a different coin.

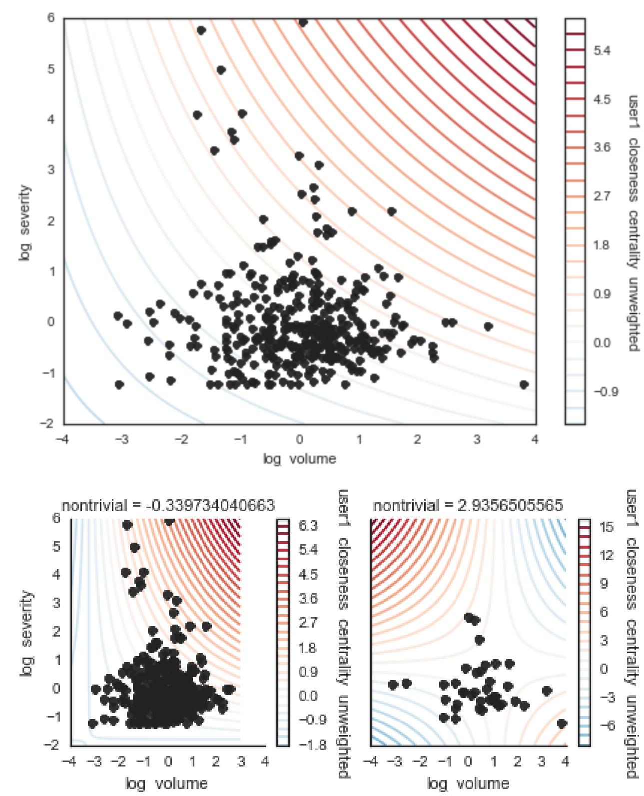
We estimate the support of our model by regularized least squares using a combination of L1 and L2 norm, with their parameters set by 5 fold cross validation (ElasticNet implementation in [**?**]) . We then estimate a OLS model over the support of the variables and calculate White robust standard errors, to allow us to examine the model coefficients and their standard errors.. Disclaimer that the regularization might make them not match (TODO: add set with normal SE that is estimated with the regularization, in results compare the coefficients) To evaluate nonlinearities and interactions in the model we fit a gradient boosted machine on the full support, cross validating its hyper parameters; as well as on the OLS selected subset.

The initial analysis pipeline and debugging, hyper-parameter setting was done using only th initial 270 of the eventual 560 in the sample. The full set of samples used for these estimates was only estimated before writing the results section, and no adjustments where made to hyper-parameters or methods after this point.

6. RESULTS

Forecasting bubbles from imidiately observable characteristics of the introducers of those coins to the forum appears to be a challenging task: of our baselines in terms of raw activity, number of members who responded and weather the code embodies substantial changes, none achieve out of sample error rates when tested in our methodology.

This is a challenging task, and models that rely on either simple activity or network metrics metrics show almost no predictive out-of-sample power, unable to explain even 1% of the variation in either task, our best models perform an order of magnitude better in both tasks. The main driver of our explanatory power is the centrality of a user in the directed network derived from the forum. This effect appears to be mediated by whether a coin involves a nontrivial technological change, the direction of the interaction reversing depending on whether it relates to magnitude or severity. Both the severity and the magnitude of bubbles increases with the centrality of the user who introduces the coins in the forums. Interestingly this effect is concentrated in different ways depending on whether the coin software is more than a trivial modification: trivial coins have more severe bubbles the more central their introducers are, while volume is greater the more central the introducer of a nontrivial coin is.



6.1 Future Work

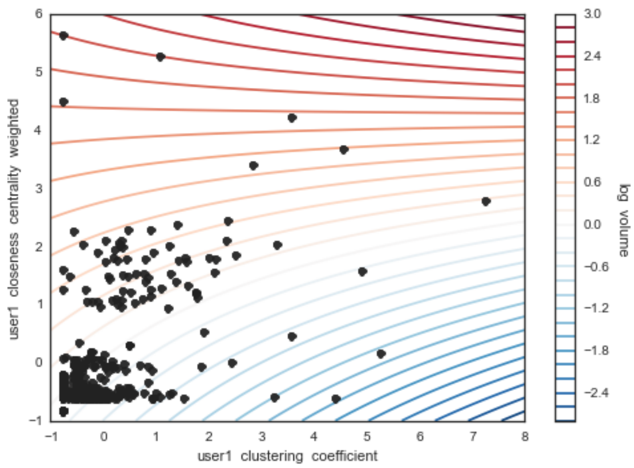
Our results suggest that bubble dynamics may be strongly influenced by founder effects, but that traditional network

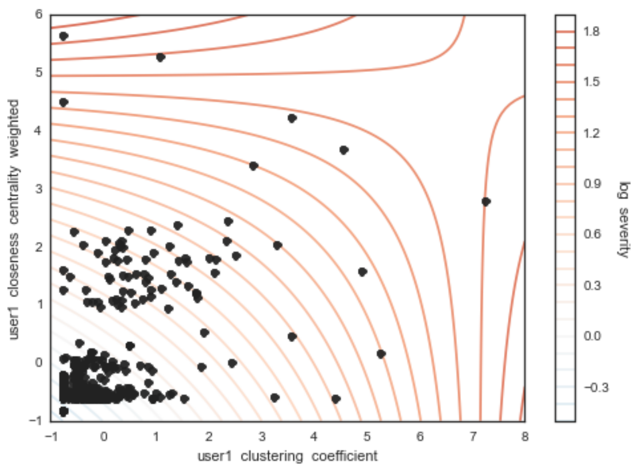
Table 1: **Severity OR**

	Model	Nontrivial	Satoshi	Network	Weighted	Interaction	All
Intercept	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)
nontrivial						0.00	0.00
						(0.00)	(0.04)
user1 betweenness centrality weighted					0.01		-0.04
					(0.07)		(0.18)
user1 betweenness centrality weighted:nontrivial						0.00	
						(0.00)	
user1 closeness centrality unweighted				0.12***		0.16***	1.69
				(0.04)		(0.05)	(3.64)
user1 closeness centrality weighted					0.15***		-1.60
					(0.04)		(3.62)
user1 closeness centrality weighted:nontrivial						0.00	
						(0.00)	
user1 clustering coefficient				0.00		0.00	0.07
				(0.00)		(0.00)	(0.05)
user1 clustering coefficient:nontrivial						0.00	
						(0.00)	
user1 days since first post	0.00			0.00	-0.03	0.00	0.02
	(0.00)			(0.00)	(0.05)	(0.00)	(0.06)
user1 degree incoming	0.03			0.02		0.04	-0.06
	(0.04)			(0.04)		(0.04)	(0.06)
user1 degree outgoing							0.04
							(0.10)
user1 degree total	0.00			0.00		0.00	-0.14**
	(0.00)			(0.00)		(0.00)	(0.06)
user1 num posts	0.00			0.00	-0.04	0.00	-0.14
	(0.00)			(0.00)	(0.08)	(0.00)	(0.11)
user1 num subjects					-0.06		0.00
					(0.06)		(0.09)
user1 pagerank weighted					0.24***		0.93***
					(0.05)		(0.17)
user1 satoshi distance							-0.34***
							(0.11)
user1 satoshi distance inf							0.30***
							(0.09)
user1 satoshi pagerank weighted							-0.64***
							(0.16)
R2	0.01	0.00	0.00	0.04	0.10	0.07	0.20
ElasticNet CV MSE:	1.00	1.00	1.02	0.98	0.95	0.97	0.91
BIC	1594	1572	1572	1585	1551	1340	1292
N	552	552	552	552	552	460	460
Adjusted-R2	-0.00	0.00	0.00	0.03	0.09	0.04	0.18
Condition Number	12.86	1.00	1.00	12.96	4.17	15.97	360428331.38

Table 2: Volume OR

	Model	Nontrivial	Satoshi	Network	Weighted	Interaction	All
Intercept	0.00	0.05	0.08*	0.00	0.07	0.07	0.04
	(0.00)	(0.05)	(0.05)	(0.00)	(0.04)	(0.04)	(0.04)
nontrivial		0.11**	0.14***	0.05	0.12***	0.11**	0.10**
		(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
user1 closeness centrality unweighted				0.16***		0.25***	0.16
				(0.05)		(0.05)	(3.50)
user1 closeness centrality weighted					0.23***		0.07
					(0.05)		(3.49)
user1 closeness centrality weighted:nontrivial						0.06	
						(0.04)	
user1 clustering coefficient				0.00		-0.07	-0.04
				(0.00)		(0.05)	(0.05)
user1 days since first post			0.08	0.00	0.05	0.06	0.02
			(0.05)	(0.00)	(0.05)	(0.05)	(0.05)
user1 degree incoming				0.00			
				(0.00)			
user1 degree outgoing				0.00	0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)	(0.00)
user1 degree total				0.00		-0.01	
				(0.00)		(0.05)	
user1 num posts				0.00	0.00		0.00
				(0.00)	(0.00)		(0.00)
user1 num subjects			0.02				
			(0.06)				
user1 pagerank weighted					-0.00		0.00
					(0.04)		(0.00)
user1 satoshi distance							0.00
							(0.00)
R2	-0.00	0.02	0.04	0.07	0.10	0.12	0.10
ElasticNet CV MSE:	1.03	0.99	0.99	0.94	0.96	0.95	0.95
BIC	1575	1292	1297	1308	1286	1283	1305
N	553	460	460	460	460	460	460
Adjusted-R2	-0.00	0.02	0.03	0.05	0.09	0.10	0.08
Condition Number	1.00	1.00	1.34	nan	3.67	3.15	200.37





measures based on their position in the aggregate discussion graph do not provide a tight characterization of this effect. As future work we are looking at coarse (positive/negative, detailed/cursory) semantic analysis of the discussion, and evidence of prior cooperation between pairs of participants in other altcoin markets, in order to attempt a more accurate characterization of core participants and their actions.

The features of the node as well as the construction of the graph are informed by a the pre-existing literature which does not allow for much complexity on the edges of the graphs beyond weight and direction. A promising avenue for future work is to explore richer models for edges, in particular allowing them to exist in latent spaces, where he influence or attnetion that is payed by responding to a user is colored by the wording of the response. Two different avenues to learn such model suggest themselves: either using higher resolution time dynamics of the prices to learn to learn a space that captures the expectations implicitly forecasted by different language, or using NLP to attempt to parse these using other external corporate to know the multidimensional valence of the words. The cross-sectional design with time separation does not allow us to take advantage of intra-coin variation. Beyond using data from forums, code repositories could also be exploited in future work as a rich source of variation beyond the binary feature of whether a coin is a trivial fork or not.

7. CONCLUSION

8. REFERENCES

[1] M. Y. Abolafia and M. Kilduff. Enacting market crisis: The social construction of a speculative bubble. *Administrative Science Quarterly*, pages 177–193, 1988.

[2] L. Bakker, W. Hare, H. Khosravi, and B. Ramadanovic. A social network model of investment behaviour in the stock market. *Physica A: Statistical Mechanics and its Applications*, 389(6):1223–1229, 2010.

[3] J. Bonneau, A. Narayanan, A. Miller, J. Clark, J. A. Kroll, and E. W. Felten. Mixcoin: Anonymity for bitcoin with accountable mixes. In *Financial Cryptography and Data Security*, pages 486–504. Springer, 2014.

[4] M. B. Clement and S. Y. Tse. Financial analyst characteristics and herding behavior in forecasting. *The Journal of finance*, 60(1):307–341, 2005.

[5] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[6] E. David and K. Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[7] P. E. Earl, T.-C. Peng, and J. Potts. Decision-rule cascades and the dynamics of speculative bubbles. *Journal of Economic Psychology*, 28(3):351–364, 2007.

[8] P. M. Garber. *Famous first bubbles: The fundamentals of early manias*. mit Press, 2001.

[9] G. Harras and D. Sornette. How to grow a bubble: A model of myopic adapting agents. *Journal of Economic Behavior & Organization*, 80(1):137–152, 2011.

[10] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.

[11] S. Moinas and S. Pouget. The bubble game: An experimental study of speculation. *Econometrica*, 81(4):1507–1539, 2013.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[13] D. M. Romero, B. Meeder, V. Barash, and J. M. Kleinberg. Maintaining ties on social media sites: The competing effects of balance, exchange, and betweenness. In *ICWSM*, 2011.

[14] D. Ron and A. Shamir. How did dread pirate roberts acquire and protect his bitcoin wealth? In *Financial Cryptography and Data Security*, pages 3–15. Springer, 2014.

[15] C. R. Shalizi. Social media as windows on the social life of the mind. In *AAAI Spring Symposium: Social Information Processing*, pages 87–92, 2008.

[16] K. Soska and N. Christin. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *Proceedings of the 24th USENIX Conference on Security Symposium*, pages 33–48. USENIX Association, 2015.
[17] P. Temin and H.-J. Voth. Riding the south sea bubble. *American Economic Review*, 94(5), 2004.

APPENDIX
Appendix A

Table 3: Volume AND

	Model	Nontrivial	Satoshi	Network	Weighted	Interaction	All
Intercept	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
nontrivial		0.15***	0.13**	0.06	0.04	0.02	0.11**
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
user1 betweenness centrality weighted					0.00		-0.04
					(0.00)		(0.08)
user1 closeness centrality unweighted				0.24***		0.18***	0.00
				(0.06)		(0.05)	(0.00)
user1 closeness centrality weighted					0.19***		0.39***
					(0.05)		(0.07)
user1 closeness centrality weighted:nontrivial						0.06	
						(0.04)	
user1 clustering coefficient				-0.04		0.00	-0.18***
				(0.06)		(0.00)	(0.06)
user1 clustering coefficient:nontrivial						0.00	
						(0.00)	
user1 days since first post			0.06	0.01	0.00	0.00	0.09
			(0.06)	(0.05)	(0.00)	(0.00)	(0.06)
user1 degree incoming			0.04	0.00	0.00	0.00	0.00
			(0.06)	(0.00)	(0.00)	(0.00)	(0.00)
user1 degree outgoing				0.00	0.00	0.00	0.09
				(0.00)	(0.00)	(0.00)	(0.11)
user1 degree total				0.00	0.00	0.00	0.08
				(0.00)	(0.00)	(0.00)	(0.12)
user1 num posts			-0.03	0.00	0.00	0.00	-0.17*
			(0.06)	(0.00)	(0.00)	(0.00)	(0.10)
user1 num subjects				0.00	0.00	0.00	-0.05
				(0.00)	(0.00)	(0.00)	(0.06)
user1 pagerank weighted					0.00		-0.17
					(0.00)		(0.19)
user1 satoshi distance							0.00
							(0.11)
user1 satoshi distance inf							-0.04
							(0.10)
user1 satoshi pagerank weighted							0.17
							(0.19)
R2	0.00	0.02	0.04	0.10	0.09	0.10	0.16
ElasticNet CV MSE:	1.01	0.99	0.99	0.94	0.95	0.95	0.94
BIC	1072	1069	1082	1078	1091	1093	1089
N	376	376	376	376	376	376	376
Adjusted-R2	0.00	0.02	0.03	0.09	0.06	0.07	0.13
Condition Number	1.00	1.00	2.05	nan	nan	nan	nan

Table 4: Severity AND

	Model	Nontrivial	Satoshi	Network	Weighted	Interaction	All
Intercept	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
nontrivial					0.00		
					(0.00)		
user1 betweenness centrality weighted							0.00
							(0.00)
user1 closeness centrality unweighted				0.00		0.00	0.15**
				(0.00)		(0.00)	(0.06)
user1 closeness centrality weighted					0.24***		0.00
					(0.05)		(0.00)
user1 closeness centrality weighted:nontrivial						0.00	
						(0.00)	
user1 clustering coefficient				0.00		0.00	0.23***
				(0.00)		(0.00)	(0.05)
user1 clustering coefficient:nontrivial						0.00	
						(0.00)	
user1 days since first post					0.00		-0.01
					(0.00)		(0.06)
user1 degree incoming				0.00	0.00	0.00	-0.09
				(0.00)	(0.00)	(0.00)	(0.07)
user1 degree outgoing							0.00
							(0.00)
user1 degree total					0.00		0.00
					(0.00)		(0.00)
user1 num posts					0.00		-0.08
					(0.00)		(0.07)
user1 num subjects							-0.01
							(0.06)
user1 pagerank weighted					0.14***		0.55***
					(0.05)		(0.18)
user1 satoshi distance							-0.11
							(0.11)
user1 satoshi distance inf							0.08
							(0.10)
user1 satoshi pagerank weighted							-0.22
							(0.17)
R2	0.00	0.00	0.00	0.00	0.18	0.00	0.27
ElasticNet CV MSE:	1.01	1.01	1.01	0.94	0.92	1.04	0.90
BIC	1072	1072	1072	1090	1040	1102	1030
N	376	376	376	376	376	376	376
Adjusted-R2	0.00	0.00	0.00	-0.01	0.16	-0.01	0.25
Condition Number	1.00	1.00	1.00	1.80	13.45	2.30	nan