Comunity Structure & Extremely Speculative Asset Dynamics

today

Abstract

Theory
ACM proceedings, LaTeX, text tagging

1 Introduction

Speculative bubbles, since at least the dutch tulip mania ([4] pp 127-31 for references) periodically take over markets. The public notoriety of Bitcoin and the massive price increases that it has seen relative to its starting prices a few years ago have lead to an explosion of attempts to create ́the next bitcoin ́ often referred to as cryptocurrencies or ́coins ́. While speculation during bubbles as a social process that has been theoretically studied [1, 3, 2, 5], data on a real social network underpinning it has not been used. The attempts to introduce these new ́coins ́ largely take place in an online forum called Bitcoin talk. We present a novel dataset based on the Bitcoin Talk forum that allows us to identify the introducers of each coin and build measures of their position in the network based on which users have engaged with whom in the forum before the coin is announced or traded (TODO clarify on when which). By considering the community structure that exists in the forum before a coin is introduced we are able to sidestep problems of reverse causation that would plague a analysis that relied on in time variation between prices and network structure. We also collect the set of prices and traded volumes across the cryptocurrencies that are introduced in the discussions on the forums, and construct measures of both the intensity (aka severity, for each dollar invested at the peak what could be recovered on average) and the magnitude (how many dollars or bitcoins where nominally traded in the asset). This allows us to evaluate the predictive power of the features of the node in our constructed network that corresponds to the user who first introduces the coin. While the magnitude of the assets traded is small relative to most financial and commodity markets, it is much larger than even the most lavishly funded experimenter could hope for. Furthermore, the rich market structure that surrounds (some 45 exchanges appear on the dataset, ranging in credibility from VC backed and registered in the US, to anonymous and mysteriously run) provides a rich source of institutional variation with extremely open data, a striking contrast to most financial or commodities market trade level data. Our contribution aims to begin in the computational social sciences a field that would have a place in the sociology of markets analogous to that of computational imaging lesion studies do in neuropsychology [1].

The evidence based uncovered by traces of their communicative interactions as they work out their thoughts about matters of common concern

quote "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."
[6]

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.[2] Since the cost of producing a 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.[3]

2 Data Description

2.1 Prices and Exchanges

Our main outcome measures are the severity of the inflation an asset price, and the magnitude of money transacted in it. We operationalize the intensity of a bubble as the proportion of a 1 dollar (TODO check currency base) that would be lost buying at the maximum price and selling after that proportionally to

[1] [, cosma2008]ff

[2] A overview of the different drug maretplaces and estimated transaction volumes can be found in [**?**]. To the best of the authors knowledge no other sector beyond speculation has even remotely subsntantial volume at present; a very primitive form of unregulated gambling Satoshi Dice, did for a brief poitnin thepast)

[3] While it is theretically possibl to engage in a distributed protocol to exchange between two cryptocurrencies, see part II of lecture 10 in [**?**]

the volume of the market till the present, we call this severity. We define the volume as the sum of the contemporaneous dollar (todo check) volume of trade. As a secondary outcome measure we consider the number of exchanges that list the coin. We scraped two major price aggregators; coinmarketcap and coininfo.

2.2 Forum Discussions

TODO: EAMAN WILL WRITE THIS

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

1. Bitcoin Discussion This is the oldest forum on the website which mainly focuses on issues only related 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 730000 posts as of July 2015, and 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 client 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 announcement 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.

TODO: discuss structure of each forum TODO: SOME preliminary statistics on forum? ave num posts per thread? TODO: NUMBER OF UNIQUE USERS IN THE COMMUNITY TODO: HOW MANY ARE ACITVE IN A 30 DAY PERIOD?

Our network measurement data consisted of discussions in Bitcoin and Altcoin online forums on bitcointalk.org from January 2010 until July 2014. Bitcoin forum is mainly concerned with Bitcoin and related technological issues whereas Altcoin discussions are concerned about alternative cryptocurrencies. Each forum consists of hundreds of threads and each thread contains many posts. Given the forum discussion data at the level of individual posts, we constructed the undirected graph of user discussions. In this network, nodes are the forum users and are connected by an edge if the corresponding users have ever interacted in any thread. The edge weights between two users are determined based on their interaction frequency. Furthermore, the weights are adjusted by size of the thread (i.e. an interaction in small threads counts more than an interaction in a large thread) and a decay factor (i.e. a recent interaction counts more than an old interaction).

We construct the network over time by replaying the posts and updating the discussion graph accordingly. Whenever a new digital currency was introduced in the forum for the first time, a snapshot of the graph was taken and used for extracting various network measures corresponding to the coin. It is important to analyze the discussion graph before the coin is mentioned for the first time to avoid any possible confounding factors. The network measures we used were as followed: number of edges, number of nodes, density, diameter, clustering coefficient, average path length and average degree. We also computed the following network measures for the user who introduced the crypto coin for the first time: clustering coefficient, closeness centrality, betweenness centrality and laplacian centrality.

3 Forum Interaction Network

TODO: EAMAN WILL WRITE THIS TODO: TALK ABOUT UNDIRECTED NETWORK WITHOUT RESULTS?

4 Analysis Variables

4.1 Prices and Exchanges

Our main outcome measures are the severity of the inflation an asset price, and the magnitude of money transacted in it. We operationalize the intensity of a bubble as the proportion of a 1 dollar (TODO check currency base) that would be lost buying at the maximum price and selling after that proportionally to the volume of the market till the present, we call this severity. We define the volume as the sum of the contemporaneous dollar (todo check) volume of trade. As a secondary outcome measure we consider the number of exchanges that list the coin.

4.2 Network Structure

TODO: EAMAN WILL WRITE THIS TODO: ENUMERATE ALL MEASURES

5 Methods Model

Initially we we start with a baseline model that considers only the user characteristics that are directly observable from their profile: the number of posts and of subjects subjects, and their seniority in the forum as measured since they first post. We then extend this model to one that uses properties that can be simply read of the relevant node in the directed graph described in section (TODO ref), the total number of edges, as well as the number of inbound and outbound. A further step is to compute network measures on the unweighted projection of the graph to a simply connected one, while alternatively one cold use the network weights (todo in ipython).

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 a classification of the literature surrounding a coin as presenting it as a potential technical advance.

We estimate linear regularized least squares using a combination of L1 and L2 norm, with their parameters set by 5 fold cross validation. We then estimate a OLS model of the support of the variables and calculate White robust standard errors, to allow for model introspection. 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. TODO add graphs showing interactions and nonlinearities; table with model comparisons.

The initial analysis pipeline and debugging, hyperparameter 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. The method will not be revised beyond this point.

6 Results

See the notebook, in particular the first 3 tables and partial plots

Table 1: Volme Prediction, symbol and name sample

	User	Node	Graph	Satoshi	weighted	technical	all
Intercept	0.00	0.00	0.00	0.00	0.00	0.21***	0.25***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.05)
technical						0.04	0.09*
						(0.05)	(0.05)
user1 closeness centrality outgoing unweighted			0.37***	0.38***	0.38***	0.25***	0.29***
			(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
user1 closeness centrality unweighted						0.00	
						(0.00)	
user1 clustering coefficient			-0.15***	-0.15***	-0.15***	0.00	0.00
			(0.06)	(0.06)	(0.06)	(0.00)	(0.00)
user1 days since first post			0.07	0.06	0.06		
			(0.06)	(0.06)	(0.06)		
user1 degree total			0.01	0.00	0.00		
			(0.07)	(0.07)	(0.07)		
user1 num posts			-0.03	-0.03	-0.03	0.00	-0.01
			(0.08)	(0.08)	(0.08)	(0.00)	(0.06)
user1 num subjects			-0.05	-0.05	-0.05		
			(0.06)	(0.06)	(0.06)		
user1 satoshi distance						0.00	-0.02
						(0.00)	(0.05)
user1 satoshi distance inf				-0.04	-0.04		
				(0.05)	(0.05)		
user1 satoshi pagerank weighted						0.00	0.00
						(0.00)	(0.07)
ElasticNet CV alpha	15.00	15.00	6.00	6.00	6.00	9.00	7.00
ElasticNet CV L1 ratio	0.10	0.10	1.00	1.00	1.00	0.50	1.00
R2	0.00	0.00	0.14	0.14	0.14	0.16	0.20
Adjusted-R2	0.00	0.00	0.13	0.13	0.13	0.14	0.18
Condition Number	1.00	1.00	3.10	3.12	3.12	8.78	2.60
ElasticNet CV MSE:	1.01	1.01	0.95	0.96	0.95	0.93	0.94
BIC	1071.97	1071.97	1050.52	1054.72	1054.72	623.01	607.68
AIC	1068.04	1068.04	1023.02	1023.29	1023.29	594.90	583.08
N	376	376	376	376	376	248	248

Table 2: volume predicion, coin name or symbol in subject

	User	Node	Graph	Satoshi	weighted	technical	all
Intercept	0.00	0.00	0.00	0.00	0.00	0.30***	0.30***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.05)
technical						0.07	0.07
						(0.05)	(0.05)
user1 closeness centrality incoming unweighted			0.04	0.04	0.02	0.00	
			(0.14)	(0.14)	(0.14)	(0.00)	
user1 closeness centrality outgoing unweighted			0.01	0.01	0.00	0.15	0.14
			(0.16)	(0.16)	(0.16)	(0.19)	(0.19)
user1 closeness centrality unweighted			0.15	0.15	0.13	0.08	0.00
			(0.23)	(0.23)	(0.22)	(0.18)	(0.00)
user1 closeness centrality weighted					0.00		0.09
					(0.00)		(0.18)
user1 clustering coefficient			-0.03	-0.03	0.00	0.00	0.00
			(0.05)	(0.05)	(0.00)	(0.00)	(0.00)
user1 days since first post			0.00	0.00	0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
user1 degree incoming						0.00	0.00
						(0.00)	(0.00)
user1 degree outgoing			0.00	0.00	0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
user1 num posts			0.00	0.00	0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
user1 num subjects						0.00	0.00
						(0.00)	(0.00)
user1 satoshi distance						0.00	0.00
						(0.00)	(0.00)
user1 satoshi distance inf						0.00	0.00
						(0.00)	(0.00)
user1 satoshi pagerank weighted						0.00	0.00
						(0.00)	(0.00)
ElasticNet CV alpha	15.00	15.00	7.00	7.00	8.00	9.00	9.00
ElasticNet CV L1 ratio	0.10	0.10	0.70	0.70	0.50	0.10	0.10
R2	-0.00	-0.00	0.06	0.06	0.05	0.15	0.15
Adjusted-R2	-0.00	-0.00	0.04	0.04	0.03	0.11	0.11
Condition Number	1.00	1.00	12.24	12.24	301.16	14.40	268.87
ElasticNet CV MSE:	1.02	1.03	0.98	0.99	0.99	0.91	0.91
BIC	1574.66	1574.66	1586.53	1586.53	1597.09	744.77	744.71

Table 3: severity, symmbol and name in title

	User	Node	Graph	Satoshi	weighted	technical	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)
user1 closeness centrality incoming unweighted			0.00	0.00	0.18**	0.00	0.00
			(0.00)	(0.00)	(0.09)	(0.00)	(0.00)
user1 closeness centrality outgoing unweighted			0.00	0.00	0.06	0.00	0.00
			(0.00)	(0.00)	(0.08)	(0.00)	(0.00)
user1 closeness centrality unweighted			0.00	0.00	0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
user1 closeness centrality weighted					0.00		0.00
					(0.00)		(0.00)
user1 clustering coefficient			0.00	0.00	0.21***	0.00	0.00
			(0.00)	(0.00)	(0.05)	(0.00)	(0.00)
user1 days since first post				0.00	-0.01		
				(0.00)	(0.06)		
user1 degree incoming			0.00		-0.06		
			(0.00)		(0.07)		
user1 degree outgoing					0.00		
					(0.00)		
user1 num posts				0.00	-0.07		
				(0.00)	(0.07)		
user1 num subjects					-0.02		
					(0.06)		
user1 pagerank weighted					0.51***		
					(0.18)		
user1 satoshi distance					-0.04		
					(0.11)		
user1 satoshi distance inf					0.08	0.00	0.00
					(0.10)	(0.00)	(0.00)
user1 satoshi pagerank weighted				0.00	-0.22		
				(0.00)	(0.17)		
ElasticNet CV alpha	15.00	15.00	10.00	10.00	5.00	10.00	10.00
ElasticNet CV L1 ratio	0.10	0.10	0.10	0.10	0.10	0.10	0.10
R2	0.00	0.00	0.00	0.00	0.28	0.00	0.00
Adjusted-R2	0.00	0.00	-0.01	-0.02	0.25	-0.01	-0.02
Condition Number	1.00	1.00	10.67	11.49	493.36	10.58	420.87
ElasticNet CV MSE:	1.01	1.01	0.92	0.93	0.90	1.04	1.05
BIC	1071.07	1071.07	1101.62	1113.48	1039.33	1101.62	1107.55

Table 4: severity, name or symbol

	User	Node	Graph	Satoshi	weighted	technical	all
Intercept	0.00	0.00	0.00	0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
user1 closeness centrality incoming unweighted			0.14***	0.00	0.15	0.17***	0.17***
			(0.04)	(0.00)	(0.20)	(0.04)	(0.04)
user1 closeness centrality outgoing unweighted					-0.33*		
					(0.17)		
user1 closeness centrality unweighted			0.00	0.00	1.13		
			(0.00)	(0.00)	(4.54)		
user1 closeness centrality weighted					-0.83		
					(4.46)		
user1 clustering coefficient			0.00		0.03		0.00
			(0.00)		(0.05)		(0.00)
user1 days since first post		0.00	0.00	0.00	-0.00		
		(0.00)	(0.00)	(0.00)	(0.05)		
user1 degree incoming		0.03	0.01	0.00	-0.09*		
		(0.04)	(0.04)	(0.00)	(0.05)		
user1 degree outgoing					0.09		
					(0.08)		
user1 degree total		0.00	0.00		-0.10***		
		(0.00)	(0.00)		(0.04)		
user1 num posts		0.00	0.00		-0.05		
		(0.00)	(0.00)		(0.08)		
user1 num subjects			0.00		-0.06		
			(0.00)		(0.06)		
user1 pagerank weighted					0.87***		
					(0.15)		
user1 satoshi distance					-0.24**		
					(0.10)		
user1 satoshi distance inf					0.26***		
					(0.09)		
user1 satoshi pagerank weighted				0.00	-0.61***		
				(0.00)	(0.14)		
ElasticNet CV alpha	15.00	9.00	9.00	10.00	0.00	8.00	8.00
ElasticNet CV L1 ratio	0.10	0.10	0.10	0.10	0.10	1.00	1.00
R2	0.00	0.01	0.05	0.00	0.16	0.05	0.05
Adjusted-R2	0.00	-0.00	0.04	-0.01	0.14	0.05	0.05
Condition Number	1.00	12.86	14.30	7.20	[illegible]	1.00	1.73

6.1 Limitations & Future Work

The features of the node in the graph we use as well as the construction ofthe graph, while informed by the literature and theory, could be substantially imporoved. We could attempt to learn the features from a more raw version of the forums, or at least learnthe parametrization of our constructor, or some larger space of weighted, or potentially labeled edges with the language used. The cross sectional design with time separation does not allow us to take advantage of intra-coin variaiton.

Beyond the forums the code repositories could be epxloited in futre work as a rich source of variation.

7 Conclusion

8 Acknowledgments

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] 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.

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

[5] 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.

[6] 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.

A Headings in Appendices