

## Research



**Cite this article:** Garcia D, Tessone CJ, Mavrodiev P, Perony N. 2014 The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *J. R. Soc. Interface* **11**: 20140623. <http://dx.doi.org/10.1098/rsif.2014.0623>

Received: 11 June 2014

Accepted: 14 July 2014

### Subject Areas:

mathematical physics

### Keywords:

social interactions, bubbles, socio-economic signals, Bitcoin

### Author for correspondence:

David Garcia

e-mail: [dgarcia@ethz.ch](mailto:dgarcia@ethz.ch)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.0623> or via <http://rsif.royalsocietypublishing.org>.

# The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy

David Garcia, Claudio J. Tessone, Pavlin Mavrodiev and Nicolas Perony

Chair of Systems Design, Department of Management, Technology and Economics, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland

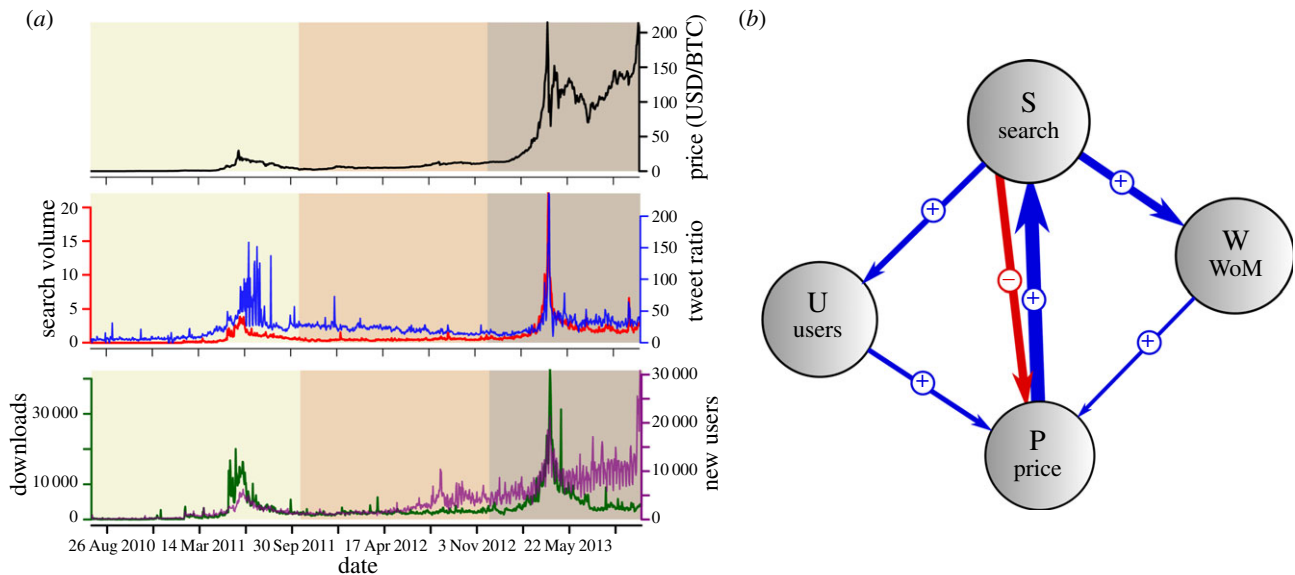
 DG, 0000-0002-2820-9151

What is the role of social interactions in the creation of price bubbles? Answering this question requires obtaining collective behavioural traces generated by the activity of a large number of actors. Digital currencies offer a unique possibility to measure socio-economic signals from such digital traces. Here, we focus on Bitcoin, the most popular cryptocurrency. Bitcoin has experienced periods of rapid increase in exchange rates (price) followed by sharp decline; we hypothesize that these fluctuations are largely driven by the interplay between different social phenomena. We thus quantify four socio-economic signals about Bitcoin from large datasets: price on online exchanges, volume of word-of-mouth communication in online social media, volume of information search and user base growth. By using vector autoregression, we identify two positive feedback loops that lead to price bubbles in the absence of exogenous stimuli: one driven by word of mouth, and the other by new Bitcoin adopters. We also observe that spikes in information search, presumably linked to external events, precede drastic price declines. Understanding the interplay between the socio-economic signals we measured can lead to applications beyond cryptocurrencies to other phenomena that leave digital footprints, such as online social network usage.

## 1. Introduction

Bitcoin [1], the best-known cryptographic currency, draws as many harbingers of imminent failure [2,3] as heralds of long-term success as a mainstream currency [4]. Throughout its 5-year existence, it has been the subject of growing attention, due in part to its rapidly increasing and very volatile exchange rate to other currencies. Amidst the hype surrounding the cryptocurrency, it is difficult to recognize which factors participate in its growth, and influence its value. Bitcoin's decentralized structure, based on the contribution of its users rather than a central authority, implies that the dynamics of its economy may be strongly driven by social factors, which are composed of interactions between the actors of the market. This paper reveals the interdependence between social signals and price in the Bitcoin economy, namely a social feedback cycle based on word-of-mouth effect and a user-driven adoption cycle.

The Bitcoin economy is indeed growing at a staggering speed: the market value of all bitcoins in circulation went from about USD \$277 K when bitcoins were first publicly traded in July of 2010, to over USD \$14B in December of 2013, hence a 50 000-fold increase in that period [5]. This came along with a stark increase in public interest, as shown by Internet search data: during the same period, the volume of Bitcoin-related searches on the Google search engine grew by over 10 000% [6]. Our hypothesis is that this growth is driven by the online actions and interactions of individual users, which leave traces of their activity. How can we use these digital traces to capture the link between market dynamics and public interest? In this paper, we provide evidence that this feedback can be understood by incorporating two types of signals: price and social information. The influence of the first was investigated through the foundational work of Fama [7] and later Grossman [8], who demonstrated that



**Figure 1.** (a) Time series of price (black) on Mt. Gox (top), number of Bitcoin-related tweets per million tweets (tweet ratio—blue), search volume (red) on Google (middle), and number of new users in the Bitcoin network (purple) and number of downloads of the bitcoin client (green) (bottom). Differently coloured backgrounds denote the three periods of our analysis. (b) Feedback diagram for the variables of our analysis. Increasing Bitcoin prices create collective attention through search volumes, which in turn triggers word of mouth about Bitcoin, leading to higher prices. A similar loop exists with the amount of users in the Bitcoin economy. Very high search volumes serve as an indicator of information search patterns before users sell their bitcoins, lowering the price. Arrows connecting variables have widths proportional to the VAR results of table 2. (Online version in colour.)

**Table 1.** Sample sizes of the Twitter, Wikipedia, Facebook, Bitcoin network and SourceForge datasets (top) and BTC exchange markets datasets (bottom).

| start date  | BTC tweets | total tweets             | Wikipedia views  |                      |
|-------------|------------|--------------------------|------------------|----------------------|
| 9 Jan 2009  | 6 827 894  | 266 306 726 448          | 6 330 676        |                      |
| end date    | users      | Facebook reshares        | client downloads |                      |
| 31 Oct 2013 | 4 717 713  | 2461                     | 3 817 506        |                      |
| market      | currency   | period                   | BTC volume       | currency volume      |
| Mt. Gox     | USD        | 17 June 2010–31 Oct 2013 | 52 273 038.8     | 1 663 743 940.58 USD |
| Mt. Gox     | EUR        | 5 Sep 2011–31 Oct 2013   | 274 870 619      | 126 206 290.65 EUR   |
| BTC-China   | CNY        | 17 June 2011–31 Oct 2013 | 2 062 919.84     | 1 383 924 703.13 CNY |
| BTC-de      | EUR        | 3 Sep 2011–31 Oct 2013   | 958 257.95       | 34 897 155.88 EUR    |

economic agents rapidly integrate common sources of information to assign a price to a good, including price itself. The role of purely social information for price formation was first studied by Bikhchandani [9], who showed that imitation is a rational strategy in periods of large volatility or in the absence of other sources of information.

In the Bitcoin economy, the fixed supply and predictable scarcity, both independent of the user base, create a strong link between public interest, user adoption and price (illustrated in the time series of figure 1a). Following from our hypothesis on the role of social interactions, a key issue is characterizing the effect of social influence [10,11] in the price variations. We quantify these socio-economic signals to provide an analytical perspective on the relationship between the Bitcoin exchange rates and the social aspects of its economy. By adopting this perspective, we reveal multiple temporal dependencies leading to the formation of Bitcoin price bubbles.

### 1.1. Four-tiered data

We use a four-tiered dataset (table 1) composed of records of exchange data, social media activity, search trends and user adoption of Bitcoin.

### 1.2. Bitcoin block chain and software client: user base

The Bitcoin block chain is a public ledger containing the full record of all public transactions in the history of the Bitcoin currency [1]. Every node of the Bitcoin network, running a Bitcoin software client, keeps a copy of the block chain. Our analysis of the block chain, as well as of the number of downloads of the software client, yields two approximations for the true number of new Bitcoin users at any time. The number of new Bitcoin users adopting the currency at time  $t$  is represented by the variable  $U_t$  in figure 1b.

### 1.3. Bitcoin exchange rates

Bitcoins (BTC) are traded for other currencies at public Internet exchanges. As of December 2013, the oldest public exchange, and the largest to trade BTC for US dollars (USD) and euros (EUR), is Mt. Gox; the largest exchange by BTC volume is BTC-China, which trades BTC for Chinese renminbi (CNY) [12]. For our analysis, we use trading data from these two exchanges and a third exchange hosted in Europe, BTC-de, including exchange rates in three currencies USD, EUR and

CNY, using BTC/USD as a historical reference. The trading price is represented by the variable  $P_t$  in figure 1*b*.

## 1.4. Information search

We measure the interest in obtaining information about Bitcoin through the normalized search volume for the term 'bitcoin' on the Google search engine. Search volume data have been shown useful to capture the information-gathering stage of the decision process of individuals, leading to insights about volume and volatility [13,14], as well as financial returns [15]. Alternatively, Wikipedia usage [16] can be used as an indicator of information-gathering.

Both indicators have been shown to lead Bitcoin prices [17], motivating the cross-validation of our results with the number of Wikipedia views for the Bitcoin page on the English Wikipedia. The search volume is represented by the variable  $S_t$  in figure 1*b*.

## 1.5. Information sharing

In our analysis, information search is a private action that need not be shared among individuals, while information sharing is strictly social. Social interaction between individuals can be measured through their level of communication. Previous works applied sentiment analysis on public messages on Twitter (tweets) to predict stock price changes [18], or linguistic patterns in instant messages to predict stock volatility [10]. In addition to sentiment, absolute online word-of-mouth levels, such as the total number of tweets or news articles, are useful in predicting price changes [19]. We measure information sharing, or online word-of-mouth communication, through the daily number of Bitcoin-related tweets  $B_t$  per million messages in our Twitter feed  $T_t$ , calculated as  $(B_t/T_t) \times 10^6$ . For additional validation, we compute an alternative by substituting the number of Bitcoin-related tweets with the number of 'reshares' of the messages posted on the oldest, regularly active public Facebook page dedicated to Bitcoin. Online word of mouth is represented by the variable  $W_t$  in figure 1*b*.

## 2. Material and methods

### 2.1. Internet datasets

We downloaded the whole Bitcoin block chain, which contains a detailed record of each block, from the website <http://blockexplorer.com> up to 5 November 2013. We retrieved search volume data from Google Trends on 5 November 2013 (<http://www.google.com/trends/explore>). We queried for the term 'bitcoin' for a set of time intervals: first the whole time period (which returned weekly volumes), then a rolling window of two months (returning daily volumes). We combined the results of these queries (see the electronic supplementary material, §S1.1, figures S1 and S2) to produce normalized daily search volumes for the whole time period. We retrieved the daily number of views of the Bitcoin page on the English Wikipedia (<http://en.wikipedia.org/wiki/Bitcoin>) by using the JSON interface of <http://stats.grok.se>. We queried each month of the dataset, getting the total amounts of views for the page each day of the queried month. We gathered the relative volume of tweets about Bitcoins through Topsy (<http://topsy.com>) on 5 November 2013, by dividing the number of tweets containing at least one of the following terms: 'BTC', '#BTC', 'bitcoin' or '#bitcoin' by the total number of tweets for each day of the study period. We extracted the number of 'reshares' of posted items

on the oldest (to the best of our knowledge), regularly active public Facebook page dedicated to Bitcoin (<http://www.facebook.com/bitcoins>). We downloaded Bitcoin market data from <http://bitcoincharts.com>, covering the largest markets for three currencies: Mt. Gox for USD, BTC-China for CNY, and Mt. Gox and BTC-de for EUR. In the main text, we use the Mt. Gox BTC/USD time series as a reference for price when not mentioned otherwise, since it is the largest market by volume throughout the study period [12].

### 2.2. Reconstructing the number of users of the Bitcoin network

We used two proxies for the number of Bitcoin users: the first is the daily number of downloads of the official Bitcoin software client from the SourceForge platform (<http://sourceforge.net/projects/bitcoin>). We also analysed the full block chain; the Bitcoin protocol is designed to preserve to a certain extent the anonymity of its users and their activity by identifying them through public keys only [1]. However, heuristics applied on transaction logs make it possible to approximately map sets of public keys to unique users [20,21]. We reconstructed the number of users of the Bitcoin network by analysing the complete set of transactions from the block chain data and applying one rule for key aggregation and one heuristic for change identification (these are described in detail in the electronic supplementary material, §S1.2).

### 2.3. Time-series stationarity

Prior to our analysis, we performed stationarity tests of the time-series  $U_t$ ,  $P_t$ ,  $S_t$  and  $W_t$ , revealing that these variables are integrated of order 1: while the time series of levels cannot be assumed to be stationary, the time series of their differences are. In the following, for each time-series  $x(t)$ , we calculate the differentiated time-series  $\Delta x(t) = x(t) - x(t-1)$ . For each of our four variables, the hypothesis of non-stationarity can be rejected at the 0.01 confidence level, and the hypothesis of stationarity cannot be confidently rejected (electronic supplementary material, table S1).

### 2.4. Vector autoregression

Since the first differences of our four variables are stationary, a vector autoregression (VAR) technique can reveal the interaction between variables. We fitted a VAR of lag 1 with a linear and a seasonal trend, measuring the time-dependent relations between the normalized changes of the four variables of our study. The significance of the relationship between variables serves as a multidimensional extension of a Granger (non-)causality test: low  $p$ -values allow us to reject the hypothesis that the changes in one variable have no linear relation to the changes of another variable in the previous day. In addition, we calculated the impulse response functions [22] of each variable to exogenous shocks on other variables, in the presence of correlated noise. We chose to combine VAR and impulse response functions to account for the multidimensional nature of our analysis and the finite sample size of our data. This provides an advantage in comparison with cross-correlation analysis between pairs of variables [22], which leads to incomplete results, as we report in the electronic supplementary material, §S3.

### 2.5. Fundamental value

It is difficult to calculate an estimate of the fundamental, or intrinsic, value of one bitcoin, which is different to its 'fair' value [23]. However, we argue that the fundamental value of one bitcoin equals at least the cost involved in its production (through mining), and therefore that we can use this cost as a lower



**Table 2.** VAR results (weights and  $p$ -values) for  $P_t$  as Mt. Gox price,  $S_t$  as Google search volume,  $U_t$  as number of downloads of the Bitcoin client and  $W_t$  as tweet ratio. Results in boldface are significant at the 0.05 level.

|              | $\Delta P_{t-1}$                       | $\Delta U_{t-1}$                       | $\Delta W_{t-1}$                        | $\Delta S_{t-1}$                        |
|--------------|--|--|---|---|
| $\Delta P_t$ | <b>0.153</b> ( $7.2 \times 10^{-8}$ )  | <b>0.137</b> ( $5.3 \times 10^{-5}$ )  | <b>0.100</b> ( $4.4 \times 10^{-4}$ )   | <b>-0.233</b> ( $2.3 \times 10^{-11}$ ) |
| $\Delta U_t$ | <b>0.184</b> ( $1.6 \times 10^{-10}$ ) | <b>-0.143</b> ( $3.0 \times 10^{-5}$ ) | -0.007 ( $7.8 \times 10^{-1}$ )         | <b>0.158</b> ( $5.8 \times 10^{-6}$ )   |
| $\Delta W_t$ | -0.032 ( $2.0 \times 10^{-1}$ )        | 8.201 ( $9.9 \times 10^{-1}$ )         | <b>-0.320</b> ( $9.4 \times 10^{-34}$ ) | <b>0.243</b> ( $1.3 \times 10^{-14}$ )  |
| $\Delta S_t$ | <b>0.386</b> ( $1.5 \times 10^{-46}$ ) | 0.016 ( $5.9 \times 10^{-1}$ )         | -0.013 ( $5.9 \times 10^{-1}$ )         | <b>0.293</b> ( $5.0 \times 10^{-20}$ )  |

bound estimate of the fundamental value. This definition has the advantage of being independent from any subjective assessment of future returns. This estimate is given by dividing the cumulated mining hash rate in a day by the number of bitcoins mined [5], to obtain the number of SHA-256 hashes needed to mine one bitcoin (this is another way to express the difficulty [1]). We then use an approximation of the power requirements for mining of 0.5 W per MHash/s, which is the average efficiency of the most common graphics processing units used to mine bitcoins [24] during our study period (mid-2010 to late 2013), and an approximation of electricity costs of  $\$0.15 \text{ KWh}^{-1}$ , which is an average of US and EU prices [25]. This yields our lower bound estimate of the fundamental value of a bitcoin, in  $\$/\text{BTC}$ .

### 3. Results

#### 3.1. Feedback loops between variables

We disentangle the feedback cycles in our system by means of a VAR [26], which captures time-dependent multidimensional linear relationships between the four variables of the analysis, with a lag of 1 day. In this statistical model, the change in each variable on a given day  $\{\Delta U_t, \Delta P_t, \Delta S_t, \Delta W_t\}$  is a linear combination of the changes in all variables  $\{\Delta U_{t-1}, \Delta P_{t-1}, \Delta S_{t-1}, \Delta W_{t-1}\}$  on the previous day, including a deterministic, a periodic and an error term [26,27]. The weight of the change of variable  $X_{t-1}$  in the equation describing the change in variable  $Y_t$  is denoted  $\phi_{X,Y}$ . Figure 1b and table 2 present a summary of these multivariate relations. The VAR reveals the following feedback cycles:

- ‘social’ cycle: search volume increases with price ( $\phi_{P,S} = 0.386$ ), word of mouth increases with search volume ( $\phi_{S,W} = 0.243$ ), and price increases with word of mouth ( $\phi_{W,P} = 0.1$ ). Simultaneously accounting for all dependencies between the four variables emphasizes the influence of word of mouth on price, revealing a stronger relation than cannot be observed with pairwise correlation analysis (more details in the electronic supplementary material, §S3). The three-way loop between  $S_t$ ,  $W_t$  and  $P_t$  represents the feedback cycle between social dynamics and price in the Bitcoin economy.
- ‘user adoption’ cycle: search volume increases with price ( $\phi_{P,S} = 0.386$ ), the number of new users increases with search interest ( $\phi_{S,U} = 0.158$ ) and price increases with increases in user adoption ( $\phi_{U,P} = 0.137$ ). This second three-way loop between  $S_t$ ,  $U_t$  and  $P_t$  models how the exchange rate of Bitcoin to other currencies depends on the number of users in the Bitcoin economy.

In addition to these two cycles, we find a negative relation from search to price ( $\phi_{S,P} = -0.233$ ). This is illustrated by a

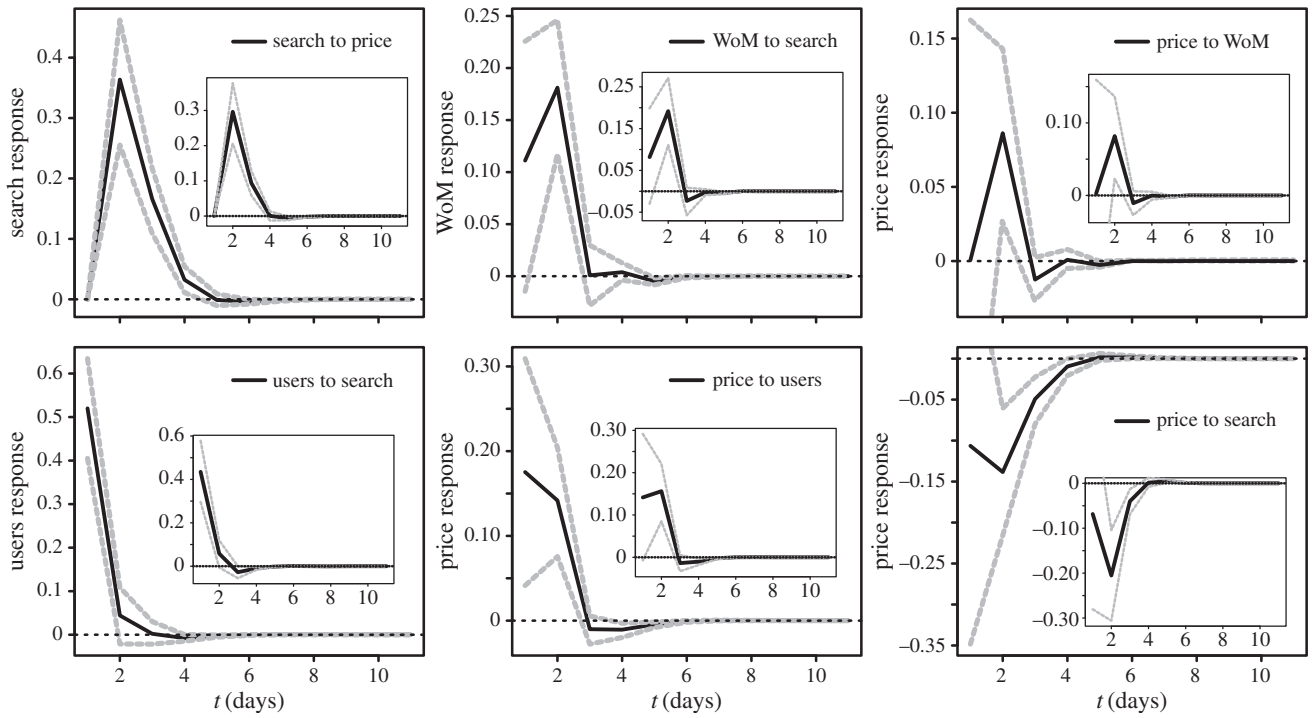
clear dyadic relation between the two variables’ extremes: three of the four largest daily price drops were preceded by the first, fourth and eighth largest increases in Google search volume the day before. The electronic supplementary material, table S2 presents the results of the same VAR with non-normalized time series; in this way, relationships between variables can effectively be quantified (e.g. an increase of 10 000 client downloads leads to an increase of  $\$3.80$  in price).

These two feedback cycles and the negative role of search in price are consistent when fitting VARs using the number of users in the network instead of client downloads as user signal  $U_t$ , Wikipedia views instead of Google search volume as search signal  $S_t$ , and Facebook Bitcoin page reshares instead of Bitcoin-related tweets as  $W_t$  (electronic supplementary material, table S3). All these results are also consistent when using other currency pairs (BTC/EUR and BTC/CNY) and data from other exchange platforms (BTC-China and BTC-de) for the analysis (electronic supplementary material, table S4).

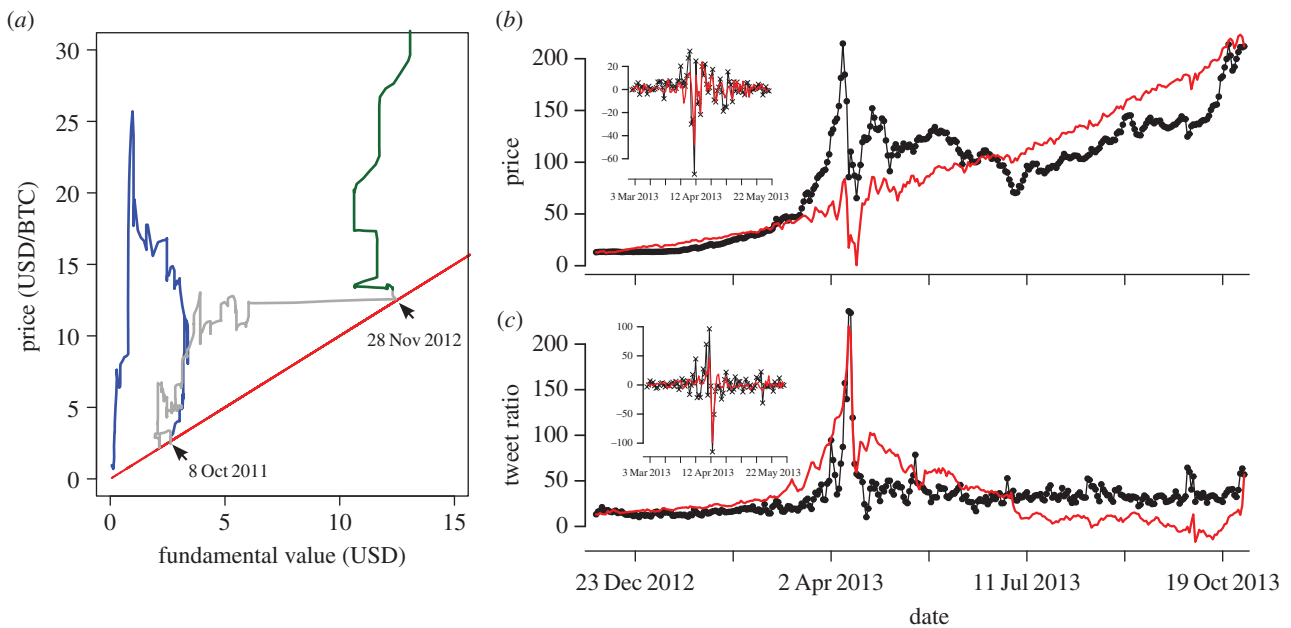
To further verify the connection between the variables in the social and user adoption cycles, we calculated the impulse response functions of the VAR results shown in table 2. The impulse response function estimates the propagation of a shock of 1 standard deviation on a variable to the other variables. To control for our finite sample size, we performed 10 000 bootstrapped estimations, correcting for correlated noise with the standard HAC method [28]. The results of the estimation for the response of variables and their 95% confidence intervals are shown in figure 2. All the pairwise relations of the feedback cycles illustrated in figure 1 are significant when analysed through impulse response functions. Search levels experience significant increases 2–4 days after price increases, and both word of mouth and number of users increase 1–2 days after strong increases in search volume. Price increases as a result of shocks on user adoption and word of mouth, and the negative influence of search in price is also evident. In contrast with findings in previous time periods [17], these response functions are similar for both Google search trends and Wikipedia page views. To further evaluate the significance of these responses, we measured the response estimate between each pair of signals for time increments of 1 and 2 days, testing for a significance of 97.5%, to achieve a 95% confidence level after Bonferroni correction. All the relations in the feedback cycles have at least one significant change, including the negative effect of search trends to price (more details in the electronic supplementary material, table S7).

#### 3.2. Reproducing price and social dynamics

The cycles presented above provide an explanation for the generation of bubbles in the Bitcoin economy. Recent findings



**Figure 2.** Impulse response function for the feedback loops between variables in our model, plus the negative relationship between search and price. Inset shows results with Wikipedia views instead of Google search volumes as an estimator for  $S$ .



**Figure 3.** (a) Trajectory of weighted mean price and mean fundamental value over weekly rolling windows between 1 Mar 2011 and 1 Mar 2013. The red line denotes strict equality between price and fundamental value. Blue dots show values before the first hitting event (18 Oct 2011), green dots show the values after the last hitting event (30 Nov 2012) and grey dots show the period in between. (b,c) Time series of the cumulative price and tweet ratio (black dots), and cumulative estimated values by the model for the period since 30 Nov 2012. Insets: time series of estimates and empirical changes of price and tweet ratio for the period between 1 Mar and 30 May 2013. (Online version in colour.)

indicate that the driving forces behind Bitcoin prices changed since its invention [29], motivating our decomposition of the study period into characteristic time windows, each of which corresponds to a distinct bubble. We do this by estimating a lower bound for the fundamental value of Bitcoin: we approximate the energy cost of producing one bitcoin, which is derived directly from the Bitcoin difficulty [1] (see Material and methods). Throughout our study period, the price stayed almost always above the fundamental value

(figure 3a: the trajectory of the weekly weighted mean price is almost exclusively on the left of the price/fundamental equality line). The trade of bitcoins at a much higher price indicates the possible presence of a bubble [30], and the events at which the market price starts diverging from the fundamental value mark the beginning of bubbles. We identify two such events: one around 18 October 2011, which marked the end of the 2011 bubble [31], and another one around 28 November 2012, which is the date at which

mining rewards halved ('Halving Day' [32]), suddenly raising the fundamental value. During the study period, the price never dropped significantly below the fundamental value, which validates our estimate for it: a drop of price below the fundamental value would introduce a paradox in which the maintenance of the public ledger by the miners is no more profitable for them. We thus define three characteristic periods, or bubbles (figure 1*b*): before the first event (P1), between the first and the second events (P2) and after the second event (P3). Running independent VAR analyses on the three periods shows that the two feedback cycles exist in the system when considering the three periods together or P3 on its own, but not if we only take into account data until 28 November 2012 (electronic supplementary material, table S5). In the following, we focus on the last period to evaluate the quality of the VAR technique in reproducing changes in the four variables in the last bubble.

The results shown in table 2, with a lag of 1 day, allow us to interpret our results in terms of feedback cycles. Nevertheless, an extended VAR with a longer lag can have larger explanatory power, despite lacking a straightforward interpretation [26,33]. By assessing the quality of the model through the Schwarz criterion [34], we find the optimal explanatory power corresponds to a lag of 4 days (electronic supplementary material, figure S4). In this best fit, the changes in each variable are calculated as linear combinations of the changes of all variables up to 4 days before. We use this optimized model to estimate daily changes in our four variables  $\Delta U_t$ ,  $\Delta P_t$ ,  $\Delta S_t$ ,  $\Delta W_t$  based on the empirical data. Adding the daily changes yields a step-wise reconstruction of the time series of each variable. Figure 3*b,c* shows the overlaid time series of estimated and empirical price and word-of-mouth levels. While this method fits daily changes in the variables, we can assess the long-term quality of the VAR by correlating their cumulative time series with the reconstructed values. We find a positive significant Pearson's correlation coefficient  $\rho$  between the observed and estimated time series for all four variables, with the price comparison yielding  $\rho = 0.8422$ , and the word-of-mouth comparison  $\rho = 0.6155$ . Additionally, reconstructed time series for the number of users ( $\rho = 0.2261$ ) and search volume ( $\rho = 0.6838$ ) are also accurate ( $\rho < 10^{-10}$ ) for all correlations. Residuals from the fit of the changes are approximately normally distributed (electronic supplementary material, figure S5), which means that within period P3 there are no structural deficiencies in our model. Finally, the VAR model correctly identifies the sign of all of the 10 largest daily price increases, and nine of the 10 largest price drops during P3 (electronic supplementary material, table S6).

## 4. Discussion

Owing to the decentralized character of the Bitcoin currency, the dynamics of its economy largely depend on the behaviour of its users, who (i) mine new bitcoins and maintain the block chain and (ii) influence the exchange rate by trading bitcoins to and from other currencies; these interactions between users form the social backbone of Bitcoin. In this paper, we used the digital traces of user activity in the Bitcoin economy to disentangle the relationships between its different variables: user base, information search, information sharing and price. While each of these four socio-economic

signals can be studied independently, the novelty of our approach lies in the combination of all of them into a robust analytical approach. Together, these signals provide a precise picture of Bitcoin's growth over the 3-year period we studied, from the first time bitcoins were publicly traded in the middle of 2010 through the successive price surges until the end of 2013. These digital activity traces [35] are a unique source of information to quantitatively describe the dynamics of large socio-economic systems [13,15]; our analysis combines for the first time social, economic and technological factors into a unified perspective on one such system.

This combined analysis reveals two positive feedback loops: a reinforcement cycle between search volume, word of mouth and price (social cycle), and a second cycle between search volume, number of new users and price (user adoption cycle). The social cycle provides evidence for interindividual influence in the decision to buy bitcoins. It translates as follows: Bitcoin's growing popularity leads to higher search volumes, which in turn result in increased social media activity on the topic of Bitcoin. More interest encourages the purchase of bitcoins by individual users, driving the prices up, which eventually feeds back on the search volumes. We hypothesize that the temporal correlation between price and search volume is mediated by the media reporting on price increases, thereby driving user curiosity and triggering their search activity. The second feedback loop we identify is the user adoption cycle, which complements the social cycle: new Bitcoin users download the client and join the transaction network after acquiring information about the technology. This growth in the user base translates to a price increase, as the number of bitcoins available for trade does not depend on demand, but rather grows linearly with time. This is a direct consequence of the deflationary nature of Bitcoin as a currency. Another important result of the VAR is the negative weight of search on price. This marks sporadic connections between large price drops and the spikes in search volume that preceded them. In other words, user search activity responds faster to negative events, such as a security breach in a Bitcoin exchange, than price. In this regard, search spikes are early indicators of price drops.

We showed the robustness of our VAR analysis, computing impulse response functions of the key couplings between variables in the feedback loops. Furthermore, we validated that our findings are not a construct of the particularities of Mt. Gox as an online exchange, reproducing our results in three different currency markets (USD, EUR and CNY). Our analysis of behavioural signals is also robust to the choice of data sources, as our results also appear when using alternative sources for search, word of mouth and user adoption. In the case of user adoption, we found that the number of new users detected after processing the block chain yields information about price changes, which is not observed when using raw Bitcoin addresses without pre-processing (see the electronic supplementary material, table S3). This indicates that identifying users from addresses is important to correctly characterize the Bitcoin ecosystem, which has been overlooked in previous studies [36].

We introduced a lower bound estimate of Bitcoin's fundamental value, based on the cost of the energy involved in mining. This allowed us to identify three characteristic time periods; by studying the three periods independently, we found that the two feedback cycles are rooted in the last period of the data (P3, end of 2012 until end of 2013), which

may either denote that the appearance of these cycles is recent, or presumably that the larger quantity of data from the last period yields a higher signal-to-noise ratio, thus helping the characterization of the cycles. This second hypothesis would also explain the more nuanced findings obtained by our analysis compared with earlier work on earlier datasets [17].

We validated the results of the VAR technique in this last period by comparing the output of the best VAR fit to the empirical time series of our four variables. It is important to note that the VAR provides a fit of the daily changes, but in this validation step we compare the cumulative levels. If the VAR errors were temporally correlated, this would create structural divergence in the VAR estimates of the levels. However, this is not what we find: the levels produced by the VAR are significantly correlated with the empirical values, and the residuals of the estimates are normally distributed (electronic supplementary material, figure S5). We also find that the dynamics of both qualitatively match (figure 3), and that the VAR correctly classifies the sign of the largest price variations (electronic supplementary material, table S6). The statistical technique we used in this paper thus proves to be a robust way of identifying the coupled dynamics of the socio-economic variables we study. It also produces accurate estimates of the future levels of any variable (including price and word of mouth) based on the past history of the system.

The two positive feedback loops we identified imply a constant increase in the price, which should drive up the Bitcoin exchange rate unsustainably. This provides an explanation for the successive periods of almost uninterrupted growth observed from mid-2010 to June 2011, and early 2012 to April 2013. It does not however explain the sudden drops observed at the end of these periods. These crashes can be attributed to external stimuli, such as the attack on the Mt. Gox platform of June 2011. The relationship we found between price drops and preceding

search spikes, while it does not explain the occurrence of the crashes, is consistent with past studies linking search volumes to price fluctuations in financial markets [14,15]. While previous work indicated that news sources do not constitute good predictors of future price changes [37], our analysis suggests that the successive price surges in the Bitcoin economy are largely due to its growing public attention. Initially, bitcoins had a negligible exchange value and were only known by a small community of technically experienced users. Our analysis suggests that the successive waves of growth of the Bitcoin economy were driven by corresponding waves of new users from public circles gradually opening to the currency. The growing user base created more exchange opportunities for Bitcoin which, at the time of writing, is accepted by a wide range of businesses, from hosting services to retail stores and illegal markets. As of December 2013, the computational power of the Bitcoin mining network sits at about  $7 \times 10^{19}$  floating-point operations per second [38]—roughly 300 times the combined power of the Top 500 supercomputers [39]. This striking figure illustrates the increasing importance of Bitcoin in the technological, social and economic landscapes, and motivates the design of policies to regulate Bitcoin usage and exchange. For such regulations to be effective, policymakers require an empirical understanding of the dynamics of Bitcoin adoption and trade. The digital traces left by the millions of users of the Bitcoin network, exchange markets, online social networks and search engines allowed us to systematically describe the dynamics of Bitcoin adoption. Our analysis of collective socio-economic signals can be applied beyond the study of cryptocurrencies, to understand other phenomena for which large amounts of data are available, such as adoption behaviour in online communities [40,41].

**Funding statement.** D.G.'s work was funded by the Swiss National Science Foundation (CR2111\_1464991). We are grateful to Frank Schweitzer for supporting this project.

## References

- Nakamoto S. 2008 Bitcoin: a peer-to-peer electronic cash system. White paper. Seattle, WA: Bitcoin Foundation.
- Hadas E. 2013 A prediction: Bitcoin is doomed to fail. *The New York Times*. See <http://dealbook.nytimes.com/2013/11/27/a-prediction-bitcoin-is-doomed-to-fail/> (accessed: 1 December 2013).
- The Blackstone Group L P. 2014 Byron Wien announces predictions for ten surprises for 2014; 2014. Press release. See <http://www.blackstone.com/news-views/press-releases/details/byron-wien-announces-predictions-for-ten-surprises-for-2014/> (accessed: 7 December 2013).
- Fontevicchia A. 2013 Winklevoss twins say Bitcoin market to hit \$400B, urge regulators not to push innovation to China. See <http://www.forbes.com/sites/afontevicchia/2013/11/12/winklevoss-twins-say-bitcoin-market-to-hit-400b-urge-regulators-not-to-push-innovation-to-china/> (accessed: 1 December 2013). Forbes.
- Data from blockchain.info. 2013. See <https://blockchain.info/charts> (accessed: 1 December 2013).
- Data from Google Trends. 2013. See <http://www.google.com/trends> (accessed: 1 December 2013).
- Fama EF, Fisher L, Jensen MC, Roll R. 1969 The adjustment of stock prices to new information. *International Economic Review* 10, 1–21. See: <http://www.jstor.org/stable/2525569>.
- Grossman SJ, Stiglitz JE. 1976 Information and competitive price systems. *Am. Econ. Rev.* 66, 246–253.
- Bikhchandani S, Hirshleifer D, Welch I. 1992 A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econ.* 100, 992–1026. (doi:10.2307/2138632)
- Saavedra S, Duch J, Uzzi B. 2011 Tracking traders' understanding of the market using e-communication data. *PLoS ONE* 6, e26705. (doi:10.1371/journal.pone.0026705)
- Lorenz J, Rauhut H, Schweitzer F, Helbing D. 2011 How social influence can undermine the wisdom of crowd effect. *Proc. Natl Acad. Sci. USA* 108, 9020–9025. (doi:10.1073/pnas.1008636108)
- Data from Bitcoin Charts. 2013. See <http://bitcoincharts.com/markets> (accessed: 7 December 2013).
- Preis T, Reith D, Stanley HE. 2010 Complex dynamics of our economic life on different scales: insights from search engine query data. *Phil. Trans. R. Soc. A* 368, 5707–5719. (doi:10.1098/rsta.2010.0284)
- Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, Weber I. 2012 Web search queries can predict stock market volumes. *PLoS ONE* 7, e40014. (doi:10.1371/journal.pone.0040014)
- Preis T, Moat HS, Stanley HE. 2013 Quantifying trading behavior in financial markets using Google Trends. *Sci. Rep.* 3, 1684. (doi:10.1038/srep01684)
- Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T. 2013 Quantifying Wikipedia usage patterns before stock market moves. *Sci. Rep.* 3, 1801. (doi:10.1038/srep01801)
- Kristoufek L. 2013 BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Sci. Rep.* 3, 3415.
- Bollen J, Mao H, Zeng XJ. 2011 Twitter mood predicts the stock market. *J. Comput. Sci.* 2, 1–8. (doi:10.1016/j.jocs.2010.12.007)
- Mao H, Counts S, Bollen J. 2011 Predicting financial markets: comparing survey, news, Twitter and search engine data. p. 10. See <http://arxiv.org/abs/1112.1051>.



20. Androulaki E, Karame G, Roeschlin M, Scherer T, Capkun S. 2012 Evaluating user privacy in Bitcoin [IACR Cryptology ePrint Archive], p. 596.
21. Reid F, Harrigan M. 2013 An analysis of anonymity in the Bitcoin system. In *Security and privacy in social networks*, pp. 197–223. Berlin, Germany: Springer.
22. Lütkepohl H. 2007 *New introduction to multiple time series analysis*. Berlin, Germany: Springer.
23. Woo D, Gordon I, Iaralov V. 2013 Bitcoin: a first assessment. Bank of America Merrill Lynch.
24. Data from the Bitcoin wiki. 2013. See <https://en.bitcoin.it/wiki/Mininghardwarecomparison> (accessed: 7 December 2013).
25. Data from the U.S. Energy Information Administration and the European Commission. 2013. See [http://www.eia.gov/electricity/monthly/epmtablegrapher.cfm?t=epmt\\_56a](http://www.eia.gov/electricity/monthly/epmtablegrapher.cfm?t=epmt_56a); [http://epp.eurostat.ec.europa.eu/statistics\\_explained/index.php/Energypricestatistics](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Energypricestatistics) (accessed: 7 December 2013).
26. Toda HY, Phillips PCB. 1994 Vector autoregression and causality: a theoretical overview and simulation study. *Econ. Rev.* **13**, 259–285. (doi:10.1080/07474939408800286)
27. Whittle P. 1953 The analysis of multiple stationary time series. *J. R. Stat. Soc. Ser. B (Methodological)* **15**, 125–139.
28. Zeileis A. 2004 Econometric computing with HC and HAC covariance matrix estimators. *J. Stat. Softw.* **11**, 1–17.
29. Kristoufek L. 2014 What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. See <http://arxiv.org/abs/1406.0268>.
30. Sornette D. 2009 *Why stock markets crash: critical events in complex financial systems*. Princeton, NJ: Princeton University Press.
31. F C. The bursting of the Bitcoin bubble. 2011. See <http://www.economist.com/blogs/babbage/2011/10/virtual-currencies> (accessed: 8 December 2013). The Economist.
32. Bitcoin Foundation. Bitcoin community celebrates 'Halving Day'. 2012. See <http://www.prlog.org/12032578-bitcoin-community-celebrates-halving-day.html> (accessed: 8 December 2013). Press release.
33. Sims CA. 1990 *Macroeconomics and reality*. Modelling Economic Series. Oxford, UK: Clarendon Press.
34. Schwarz G. 1978 Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464. (doi:10.1214/aos/1176344136)
35. Lazer D *et al.* 2009 Computational social science. *Science* **323**, 721–723. (doi:10.1126/science.1167742)
36. Kondor D, Pósfai M, Csabai I, Vattay G. 2014 Do the rich get richer? An empirical analysis of the Bitcoin transaction network. *PLoS ONE* **9**, e86197. (doi:10.1371/journal.pone.0086197)
37. Cutler DM, Poterba JM, Summers LH. 1989 What moves stock prices? National Bureau of Economic Research. 2538. See: <http://www.nber.org/papers/w2538>.
38. Bitcoin Watch. See <http://www.bitcoinwatch.com/> (accessed 2 December 2013).
39. Top500.org. See <http://www.top500.org/lists/2013/11/> (accessed 2 December 2013).
40. Onnela JP, Reed-Tsochas F. 2010 Spontaneous emergence of social influence in online systems. *Proc. Natl Acad. Sci. USA* **107**, 18 375–18 380. (doi:10.1073/pnas.0914572107)
41. Garcia D, Mavrodiev P, Schweitzer F. 2013 Social resilience in online communities: the autopsy of friendster. In *Proc. First ACM Conf. on Online Social Networks*. COSN '13; 2013. pp. 39–50.