RESEARCH

The Butterfly "Affect": Impact of Development Practices on Cryptocurrency Prices

Silvia Bartolucci[1,5*], Giuseppe Destefanis[2,5], Marco Ortu[3], Nicola Uras[3], Michele Marchesi[4] and Roberto Tonelli[4]

[*] Correspondence: s.bartolucci@imperial.ac.uk
[1] Imperial College Business School, London, UK
Full list of author information is available at the end of the article

Abstract

The network of developers in distributed ledgers and blockchains open source projects is essential to maintaining the platform: understanding the structure of their exchanges, analysing their activity and its quality (e.g. issues resolution times, politeness in comments) is important to determine how "healthy" and efficient a project is. The quality of a project affects the trust in the platform, and therefore the value of the digital tokens exchanged over it.

In this paper, we investigate whether developers' emotions can effectively provide insights that can improve the prediction of the price of tokens. We consider developers' comments and activity for two major blockchain projects, namely Ethereum and Bitcoin, extracted from Github. We measure sentiment and emotions (joy, love, anger, etc.) of the developers' comments over time, and test the corresponding time series (i.e. the *affect time series*) for correlations and causality with the Bitcoin/Ethereum time series of prices. Our analysis shows the existence of a Granger-causality between the time series of developers' emotions and Bitcoin/Ethereum price. Moreover, using an artificial recurrent neural network (LSTM), we can show that the Root Mean Square Error (RMSE) – associated with the prediction of the prices of cryptocurrencies – significantly decreases when including the affect time series.

Keywords: blockchain; open source software development; data mining.

1 Introduction

The ecosystem of cryptocurrencies traded and exchanged every day has been exponentially growing over the past ten years. The platforms – distributed ledgers and blockchains – cryptocurrencies rely upon to be created and transferred are developed (in most cases) in the form of open source projects. Developers from across the globe are constantly contributing to open source projects maintaining the codes and software that ensure the platform's correct functioning. According to a Deloitte report[1], there are currently more than 6500 active projects connected to distributed ledger technologies (DLT) and blockchains. The pioneering ones are the

[1] https://www2.deloitte.com/content/dam/insights/us/articles/4600_Blockchain-five-vectors/DI_Blockchain-five-vectors.pdf

Bitcoin and Ethereum projects, whose associated tokens also dominate the crypto scene by market capitalisation.

Investors are attracted to this technology, not only by its future outlooks and potential but also by the excess returns on their investments that can be achieved by exploiting the highly volatile crypto-market. Nonetheless, valuation and pricing of cryptocurrencies and digitally native tokens remains a non-trivial task, due to the peculiarity of the platforms, users and investors in the space.

Quantitative investigations aimed at extracting information from the time series of cryptocurrency prices and predicting prices drivers [1] or the next most likely jump, range from theoretical models of pricing and adoption of digital tokens [2, 3, 4] to machine learning [5, 6] and neural network-driven [7] forecasts of prices and returns. Analyses of the cryptocurrency markets [8, 9, 10] yielded insights on their maturity, efficiency and structure. A large body of literature is also looking at the volatility of cryptocurrencies, from the model estimation point of view [11, 12] as well as by extrapolating the mechanisms driving the fluctuations. Studies showed, for example, a strong correlation with global economic activity [13, 14] and volume of trades [15].

In the crypto space, where everything is decentralised and shared in a peer-to-peer fashion between users, developers and investors, "social" aspects appear to play a crucial role: discussions about platforms' quality are held over public forums (e.g. Reddit), news about next developments are shared over the informal news channel of Twitter and updates on development activities are publicly accessible over open source development platforms such as Github. Investors' sentiment and trading activities, which in turn impact prices, are, therefore, inevitably informed and influenced via those channels. For this reason, new types of data have been recently used to improve models and predictions. "Social" sentiment is extracted using data gathered from users' online communities [16] – e.g. online forums such as BitcoinTalk[2] – and from online news and tweets [17, 18, 19]. For instance, a suitably built sentiment index can be used to test for speculative bubbles in cryptocurrency prices [20]. More broadly, Google search data related to cryptocurrencies can be relevant to characterise the set of Bitcoin users [21]. Temporal topic analysis of Bitcoin and Ethereum discussions on Reddit also show correlations with variations of cryptocurrency prices [22].

Developers in open source blockchain and DLTs projects are also crucial entities, responsible for the maintenance and updates of the platforms. The idea that the human aspects of software development are of paramount importance to ensure high team productivity, software quality and developers satisfaction is already well-established in software engineering [23, 24, 25]. These studies have shed light on the importance of all the social and human aspects associated with the software development processes, and empirically demonstrated how a positive environment may have an impact on team productivity, software quality and developers' satisfaction [26, 27]. Moreover, standard metrics extracted from open source development platforms such as Github[3] and Bitbucket[4] can be used to rank the top crypto tokens [28]: metrics include number of commits and issues, forks and number of contributors to the code.

[2] https://bitcointalk.org

[3] https://github.com

[4] https://bitbucket.org

Tools to extract sentiment specifically built for the software engineering domain and language are also available. For example, Murgia et al. [25] demonstrated the feasibility of a machine learning classifier using emotion-driving words and technical terms to identify developers' comments containing gratitude, joy and sadness. Islam et al. [29] studied how developers' emotions affect the software development process. They reconstructed the emotional variations by means of a sentiment analysis on commit messages using the SentiStrength tool [30], showing how emotions influence different typologies of software development tasks.

In this paper, we focus our investigation precisely on the impact of developers' activities and emotions, sentiment and politeness on the cryptocurrencies issued and transferred over the platform they contribute to develop. In particular, we consider comments written by GitHub contributors of the two main blockchain projects, Bitcoin and Ethereum, and we perform *emotions mining* (love, joy, anger, sadness), *sentiment analysis* [25], *politeness and VAD analysis*[5] of the comments [24, 31]. In the following, we will generally refer to emotions, sentiment, politeness and VAD metrics as *affect metrics*, in line with recent works in psychology and computer science (e.g. [32, 33]), where *affect* is an umbrella term for discrete emotional states as well as emotional dimensions and moods. In Sec. 2, we will describe in more details the meaning of the affect metrics and how they are measured.

The main idea of this study is to understand whether emotions mining, sentiment analysis, politeness, and VAD analysis can be used to improve the prediction power of machine learning algorithms for the returns of the Bitcoin/Ethereum cryptocurrency. More generally, these metrics could be useful to monitor the health and quality of projects and platforms from a software engineering point of view.

We aim at understanding the interplay between developers' affect and cryptocurrency returns and we will focus on the two following aspects:

- Does the affect of Bitcoin and Ethereum communities influence variations in returns?
  Using Granger causality tests we will show that the affect metrics extracted from the contributors' comments influence the cryptocurrency prices.
- Is the affect of Bitcoin and Ethereum communities able to improve the error on the prediction of returns?
  Using a LSTM neural network we will show that including the affect time series as features in the training set significantly improves the prediction error.

This paper is organised as follows. In Sec. 2, we describe the dataset, the process to construct the affect time series and the tools used for the analyses (Granger causality test and Long-Short term memory for the prediction of returns). In Sec. 3, we present the results and their implications. In Sec. 4, we discuss the limitations of this study. Finally, in Sec. 5 we summarise the main findings.

2 Dataset and Methods

In this section, we describe how affect time series are constructed using the comments of Ethereum and Bitcoin developers on Github for the period of December 2010 to August 2017.

[5] Valence, Arousal and Dominance: these metrics are used to respectively evaluate the (i) engagement, (ii) confidence and (iii) responsiveness of a person in conducting a task or an activity. More details will follow in Sec. 2.1.

Both the Bitcoin and Ethereum projects are open source, hence the code and all the interactions among contributors are publicly available on GitHub [34]. Active contributors are continuously opening, commenting on, and closing so-called "issues". An issue is an element of the development process, which carries information about discovered bugs, suggestions on new functionalities to be implemented in the code, or new features actually being developed. Monitoring the issues constitutes an elegant and efficient way of tracking all the phases of the development process, even in complicated and large-scale projects with a large number of remote developers involved. An issue can be "commented" on, meaning that developers can start subdiscussions around it. They normally add comments to a given issue to highlight the actions being undertaken or to provide suggestions on its possible resolution. Each comment posted on GitHub is timestamped, hence it is possible to obtain the exact time and date and generate a time series for each affect metric considered in this study.

An example of a developer's comment extracted from Github for Ethereum can be seen in Table 1. Quantitative measures of sentiment and emotions associated with the comments, as reported in this example, are computed using state-of-the-art tools of textual analysis (further details below). The affect metrics computed for each comment are emotions such as love (L), joy (J), anger (A), sadness (S), VAD (valence (Val), dominance (Dom), arousal (Ar)), politeness and sentiment (Pol and Sent respectively).

Comment	L	J	A	S	Val	Dom	Ar	Pol	Sent
Perhaps there's simply nothing new to translate? The reason I updated Transifex in the first place was to be sure the strings with subtle English changes (that don't change the meaning) didn't reset the translation - so those were imported from the old translations. Though I seem to recall at least one truly new string - Transaction or such.	0	0	0	1	1.93	1.88	1.26	imp	1

Table 1: Example of comments and the corresponding values of affect (love (L), joy (J), anger (A), sadness (S)), VAD (valence (Val), dominance (Dom), arousal (Ar)), politeness and sentiment (Pol and Sent respectively).

The Bitcoin and Ethereum price time series were extracted from the API of CoinMarketCap[6] using daily closing prices.

2.1 Measuring Affects Metrics

In our analysis, we focus on four main classes of affect metrics: emotions (love, joy, anger, sadness), VAD (valence, arousal, dominance), Politeness, Sentiment. As we specify below, for each affect metric class, we use a tailor-made tool to extract it from the text of the comments.

For the detection of emotions, we use the tool developed by Ortu et al. [35] and extended by Murgia et al. [36]. This tool is particularly suited for our analysis as

[6] https://coinmarketcap.com/

the algorithm has been trained on developers' comments extracted from Apache, a Jira-based data repository, hence within the Software Engineering domain. The classifier is able to detect love, anger, joy and sadness with an F_1 score[7] close to 0.8 for all of them.

Valence, Arousal and Dominance (VAD) represent conceptualised affective dimensions that respectively describe the interest, alertness and control a subject feels in response to a certain stimulus. In the context of software development, VAD measures may give an indication of the involvement of a developer in a project as well as their confidence and responsiveness in completing tasks. Warriner et al.'s [37] has created a reference lexicon containing 14,000 English words with VAD scores for Valence, Arousal, and Dominance, that can be used to train the classifier, similarly to the approach by Mantyla et al. [31]. In [31], the authors extracted valence-arousal-dominance (VAD) metrics from 700,000 Jira issue reports containing over 2,000,000 comments. They showed that issue reports of different type (e.g., feature request vs bug) had a fair variation in terms of valence, while an increase in issue priority would typically increase arousal.

For politeness detection, we use the tool proposed by Danescu et al. [38], which output a binary classification of the text as *polite* or *impolite*. This tool is particularly suitable in the context of our analysis as the algorithm has been trained using over 10,000 manually labelled requests from Wikipedia and StackOverflow. Indeed, in both data sources–but more specifically StackOverflow–contributors make use of technical terms and jargon, similarly to conversations among developers in online forum or development platforms.

Finally, the sentiment is measured using *Senti4SD* tool [39]. The algorithm extracts the degree of positive (ranging from 1 to 5), neutral (0) and negative (ranging from -1 to -5) sentiment in short texts. This tool is also trained on developers' comments.

2.2 Affect Time Series

Once numerical values of the affect metrics are computed for all comments (as shown in the example in Table 1), we consider the timestamps (i.e. dates when the comments were posted) to build the corresponding affect time series. The affect time series are constructed by aggregating sentiment and emotions of multiple comments published on the same day. For a given affect metric, e.g. anger, for a specific day, we construct the time series by averaging the values of the affect metric over all comments posted on the same day.

In Table 2 and 3 we report the summary statistics of the affect time series for Bitcoin and Ethereum respectively. Plots of all the affect time series concerning Bitcoin and Ethereum are also available within the Supplementary Material (see Figure 1, 2).

In Figure 1, we also show the boxplots of the data distributions for each affect time series for the Bitcoin and Ethereum cases.

The box width gives an indication of the sample's variability. In the Bitcoin case, all the affect metrics show a small variance, particularly if we consider *anger*, *joy*

[7] The F_1 score tests the accuracy of a classifier and it is calculated as the harmonic mean of precision and recall.

Statistic	sentiment	arousal	valence	dominance	anger	sadness	joy	love
mean	-0.043081	0.984999	1.431276	1.442303	0.040907	0.244030	0.077308	0.049961
std	0.807408	0.541344	0.821061	0.805969	0.227526	0.595813	0.291157	0.227564
min	-4.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.750000	1.080000	1.100000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	1.040000	1.470000	1.490000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	1.270000	1.840000	1.860000	0.000000	0.000000	0.000000	0.000000
max	4.000000	6.150000	7.890000	7.220000	12.000000	38.000000	12.000000	4.000000

Table 2: **Summary statistics of affect metrics for Bitcoin.** Mean, standard deviation, min-max values considering *all* Github comments for sentiment, arousal, valence, dominance, anger, joy, love.

Statistic	sentiment	arousal	valence	dominance	anger	sadness	joy	love
mean	0.039835	1.271606	1.831578	1.829091	0.035780	0.224817	0.056057	0.132588
std	0.794382	0.837046	1.264279	1.188470	0.217238	0.547257	0.273674	0.365146
min	-3.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.900000	1.290000	1.300000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	1.150000	1.640000	1.660000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	1.420000	2.050000	2.060000	0.000000	0.000000	0.000000	0.000000
max	4.000000	5.570000	8.210000	7.000000	6.000000	15.000000	14.000000	3.000000

Table 3: **Summary statistics of affect metrics for Ethereum.** Mean, standard deviation, min-max values considering *all* Github comments for sentiment, arousal, valence, dominance, anger, joy, love.

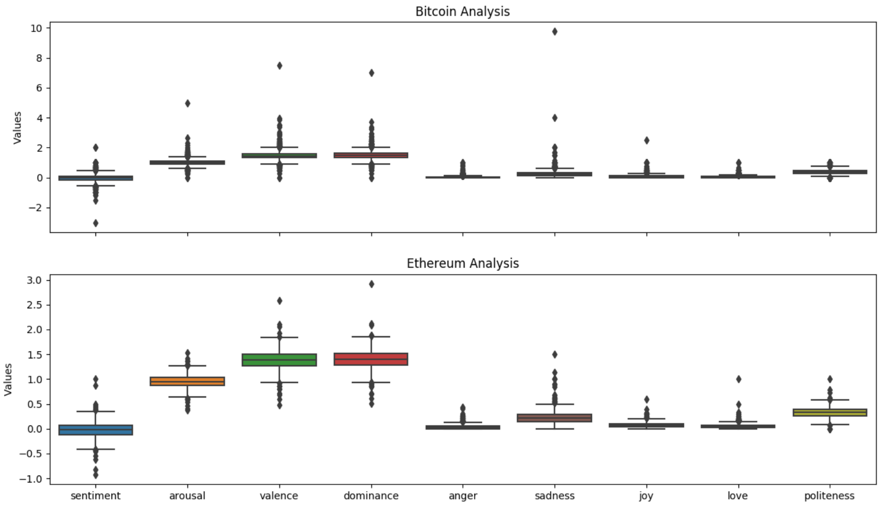


Figure 1: **Distributions of the affect time series.** Boxplot of the Bitcoin (top panel) and Ethereum (bottom panel) distributions for all affect metrics and all Github comments.

and *love* time series. Moreover, all distributions are symmetric, except that for the *anger*, *joy*, *sadness* and *love* samples. For Ethereum, instead, the time series of sentiment, arousal, valence and dominance present a broader distribution compared to the corresponding Bitcoin ones. Further analyses of the stationarity of the time series can be found in the Supplementary Material (Sec. 1).

2.3 Granger Causality test

The Granger causality test is a statistical hypothesis test useful to assess whether a given time series shows some potential predictability power on another time series. In the *Granger-sense*, a time series X *Granger-causes* Y if X is able to improve the prediction of Y with respect to a forecast, considering only past values of Y [40]. Equivalently, if we define Γ'_τ as the information set of the form $(x_t, ..., x_{t-\tau}, y_{t-1}, ..., y_{t-\tau})$ (where τ is the number of lags or observations included in the regression), then x_t *Granger-causes* y_t if the variance of the optimal linear predictor of y_t based on Γ'_τ is smaller than the variance of the optimal linear predictor of y_t based on the information set $\Gamma_\tau = (y_{t-1}, ..., y_{t-\tau})$ of lagged values of y_t only:

$$\sigma^2(y_t|y_{t-\tau}, x_{t-\tau}) < \sigma^2(y_t|y_{t-\tau}), \forall\tau \in \mathbf{N} \ . \tag{1}$$

The procedure of the *Granger-causality* test is as follows.

1 The time series of cryptocurrency returns (e.g. BTC returns) (y_t) is regressed on its past values excluding the metric series in the regressors. The so-called *restricted* regression can be written as

$$y_t = \alpha + \sum_{i=1}^{\tau} \rho_i y_{t-i} + \xi_t \ , \tag{2}$$

where α is a constant and the error term ξ_t is an uncorrelated white-noise process. We, then, calculate the restricted sum of squared residuals (SSR_r)

$$SSR_r(\tau) = \sum_{i=1}^{N} \left[y_i - \hat{y}_i(\Gamma_\tau)\right]^2 \ , \tag{3}$$

where N is the number of observations, τ is the number of lags included in the regression, Γ_τ is the information set, and $\hat{y}_i$ are the predicted values.

2 We compute a second regression including the lagged values of the affect time series in the regressors. This *unrestricted* regression reads

$$y_t = \alpha + \sum_{i=1}^{\tau} \rho_i y_{t-i} + \sum_{i=1}^{\tau} \gamma_i x_{t-i} + \xi'_t \ . \tag{4}$$

As before, we evaluate the unrestricted sum of squared residuals (SSR_u) as follows

$$SSR_u(\tau) = \sum_{i=1}^{N} \left[y_i - \hat{y}_i(\Gamma'_\tau)\right]^2 \ . \tag{5}$$

3 Finally, if $SSR_u < SSR_r$ the affect time series considered for the analysis *Granger-causes* the cryptocurrency returns series.

To determine the presence of (direct and reverse) Granger causality between affect and return time series we use a two-step approach: (i) we first tested the null-hypothesis rejecting it if the p-values are below the chosen significance level and then

(ii) we restricted the set of time series to the ones minimising also the information loss (using the Akaike criterion specified below). Both approaches are standard tools used for the optimal lag selection in the econometric literature [41, 42, 43]. For the sake of completeness, and since in the literature issues and biases in lag estimation have been extensively discussed, in the Supplementary Material we also provide an independent estimation of the time lag parameter using the Bayesian Information Criterion (BIC) [44], which for instance is known to be biased in favour of more parsimonious models in terms of number parameters [41]. Nonetheless, in general, the AIC criterion appears to perform better when compared to other tests [41, 45].

It is worth highlighting that due to the different search procedures employed by the various methods, we should expect different lag lengths being deemed optimal [46]. Nonetheless, the main goal of our analysis is not to precisely estimate the time lag for the causality – which would be an unrealistic task – but rather to demonstrate via independent statistical tests that a non-spurious causality relationship between the affect and the return time series does exist.

For this analysis we have implemented the *grangercausalitytest* test using the *statsmodels* Python library [47]. This tool tests for *Granger non-causality* of two time series, i.e. the null hypothesis H_0 is that the chosen affect metric series does not *Granger-cause* the Bitcoin or Ethereum returns series. We reject the null hypothesis if the p-values are below a desired size of the test, choosing a 5% significance level. The p-values are computed using the Wald test as per the standard Python *statsmodel* libraries [48, 47]. The number of lags included in the regression models can be tuned by the τ parameter. For any fixed value of τ, a *Granger causality* test is computed for all lags up to τ.

The two possible outcomes of the Granger test are:

- The observed p-values are less than the 5% significance level: rejection of the null hypothesis H_0. The affect time series *Granger cause* the cryptocurrency returns one.
- The observed p-values are greater than the 5% significance level: H_0 cannot be rejected. The affect time series does not *Granger cause* the cryptocurrency returns one.

In presence of significant causality between returns and affect time series, then the *AIC* metric is monitored for the two models (direct and reverse causality) for each lag value to check for consistency with the results obtained via the *Granger causality* test. The *Akaike Information Criterion* (*AIC*) is a statistical tool, based on information theory, that can be used for model selection.

The *AIC* metric provides an estimate of the quality of a given model, based on the loss of information: the best model minimises the information loss. *AIC* for least squares model fitting can be mathematically defined as

$$AIC = 2(k+1) + n\log(SSR)\ , \tag{6}$$

where n is the sample size and k is the number of parameters [49]. We then look for the lag value for which the *AIC* is minimal. If this predicted value is compatible with the lag estimated using the p-values, we can further corroborate that the Granger causality test has not highlighted a spurious, non-statistically significant correlation. Therefore, we restrict the set of affect time series effectively showing Granger causality with returns, to the ones for which not only the p-value is below the chosen significance level but also the AIC is minimal.

We perform the Granger test on the stationary affect time series selected via the analysis available in Sec. 1 of the Supplementary Material. According to our analysis, the stationary affect time series that we will consider for the Bitcoin case are *sentiment*, *sadness*, *arousal*, *valence*, *love* and *dominance*. For the Ethereum case we will use *sentiment*, *anger*, *arousal*, *valence*, *love*, *dominance*, *joy* and *politeness*.

It is worth noting that the Granger causality test is sensitive to the number of lags input in the model. For this reason, we have analysed a large range of lags, of the order of five months. More specifically, the τ parameter was set to 150.

2.4 Long Short-Term Memories and predictions

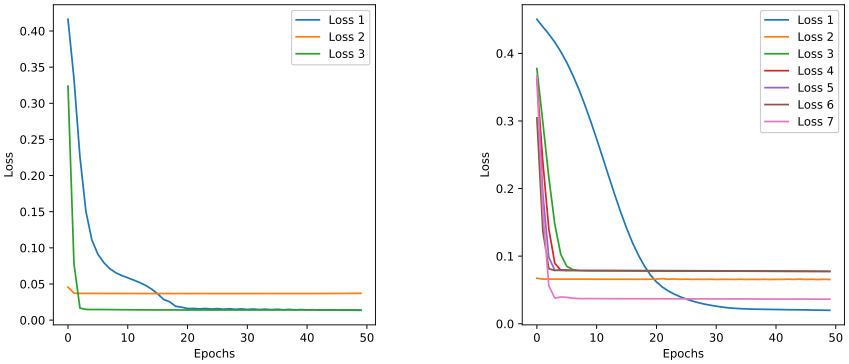
For our prediction task of the cryptocurrency prices, we use a Recurrent Neural Network (RNN). A RNN, at its most fundamental level, is simply a type of densely connected neural network. However, the key difference with respect to normal feed-forward networks is the introduction of time, with the output of the hidden layer in a recurrent neural network being fed back into itself. RNNs are often used in stock-market predictions [50, 51, 52] and more recently also for Bitcoin and cryptocurrency prices [53, 52].

In this analysis, we use a Long-Short Term Memory (LSTM) RNN to predict Bitcoin and Ethereum returns. In our model, we use the previous day returns and affect metrics for the prediction of the returns of the current day (1-day forecast horizon). We decided to use this short forecast horizon model – which is normally a benchmark of more sophisticated prediction algorithms – as we are mostly concerned about demonstrating a possible improvement in Root Mean Square Error (RMSE)[8] when inputting in the model the affect time series rather than building a sophisticated prediction model.

The affect time series used for this analysis are the ones that showed Granger-causality with the Bitcoin and Ethereum returns time series. Indeed, the test assessed whether a given affect time series had some potential predictive power over the cryptocurrency returns time series. As reported in Sec. 3.1, we selected *sentiment* and *sadness* for Bitcoin and *sentiment*, *anger*, *arousal*, *dominance*, *valence* and *love* for Ethereum.

We designed the LSTM with 50 neurons in the first hidden layer and 1 neuron in the output to predict the cryptocurrency returns. To configure the LSTM, we use a sigmoid activation function, we calculate the *Mean Absolute Error* (MAE) loss function and we use the efficient Adam version of stochastic gradient descent [54] for the optimal choice of models' parameters. We train the LSTM, first, using only data related to the cryptocurrency (Ethereum or Bitcoin) returns time series and, then, we incrementally add the correlated (via Granger-causality) affect metrics features.

[8] The RMSE is defined as the standard deviation of the residuals or prediction errors, i.e. $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$, where n is the number of observation, $y_i, i = 1, \ldots, n$ are the observed values and $\hat{y}_i, i = 1, \ldots, n$, the predictions.



(a) Bitcoin

(b) Ethereum

Figure 2: **RNN Loss against number of epochs.** Panel (a) RNN trained with Bitcoin only data (1) first and then sequentially adding sentiment (2) and sadness (3).
Panel (b) RNN trained with Ethereum returns data only (1), then adding sequentially sentiment (2), anger (3), arousal (4), valence (5), dominance (6), love (7).

We first apply the LSTM using only the cryptocurrency returns (Bitcoin or Ethereum returns time series) as a feature, i.e. solving in this case a univariate regression problem. Then, we incrementally add the affect metrics, i.e. considering a multivariate regression problem, to analyse potential effects on the RMSE associated with the predictions. Our analysis is performed by training the LSTM for 50 epochs and recording for each epoch the corresponding RMSE value. Figures 2a and 2b show the loss of the RNN models against the epochs. We can see that after 50 epochs the loss converges to a stationary value for all models. Finally, all models where trained using 70% of data for training and 30% for testing.

3 Results

In this section, we summarise the results of our analysis concerning testing for (i) causality between affect time series and cryptocurrency returns and (ii) improvement in Root Mean Square Error (RMSE) for the prediction of returns when including affect time series.

3.1 Does the affect of Bitcoin and Ethereum communities influence variations in returns?

In this section, we focus on understanding if there exists a causal relationship between affect time series and the time series of Bitcoin/Ethereum returns. The analysis is performed using the *Granger causality* test [40], which informs on whether changes in a time series – in our case the returns time series – are induced or connected to a variation in a second correlated time series – in our case the affect time series. Details on the Granger test can be found in Sec. 2.3. As we will show in the following, the Granger test is detecting significant Granger-causality (both direct and reverse causality) between affect time series and cryptocurrency returns. Via this analysis, we are also able to give an estimate of the time lag or delay after which effects of variations in the affect time series are "visible" in the cryptocurrency returns time series.

3.1.1 Granger causality test - Bitcoin

Let us start with the Bitcoin returns time series analysis. The Granger test highlights that only *sentiment* and *sadness* metrics *Granger-cause* the Bitcoin series. According to the test, instead, there is no causal relationship between the Bitcoin returns and *arousal*, *valence*, *love* and *dominance* time series, for any considered lag value. In order to select the time lag for the Granger causality, we monitor the p-values as a function of the time lag and select – among the time lags with p-values falling below the significance level – the time lag associated with the minimal p-value.

As an illustration, we show in Fig. 3 the p-values obtained for each lag value, up to the chosen τ for the two affect time series that displayed statistically significant Granger causality with the Bitcoin returns time series.

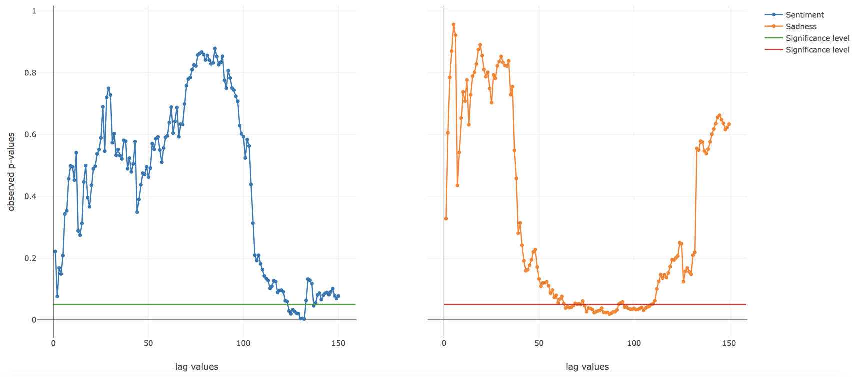


Figure 3: **Bitcoin direct causality - p-values as a function of lags.** Left: Bitcoin returns - sadness time series direct causality. Right: Bitcoin returns - sadness time series direct causality.

A *reverse Granger-causality test* was also conducted, in order to test whether cryptocurrency prices influence the affect time series, hence developers' behaviour and feelings. Specifically, we obtain that Bitcoin returns Granger-cause only *sentiment* and *sadness* affect metrics. In this case we therefore deal with a *bidirectional causality*, whereby *sadness* and *sentiment* series increase the prediction of the Bitcoin price returns and vice versa. Table 4 contains the minimal p-values and associated time lags for *all* affect metrics for direct and reverse causality.

We have also checked the *AIC* values of the models to ascertain that the Granger test was not capturing spurious effects. *AIC* values as a function of the time lags can be found for *all* affect metrics Granger-causing the Bitcoin returns in Sec. 2 of the Supplementary Material. An estimation of the time lag using a different information criterion, namely the BIC is also provided in the Supplementary Material (Sec. 2).

Table 4: **Bitcoin Granger Causality tests.** Minimum observed p-values and corresponding *lag values* for sadness and sentiment times series. We highlight in green the cases showing significant (direct or reverse) Granger-causality.

	Granger Causality Test		*Reverse Granger Causality Test*	
Series	**p-value**	**lag value**	**p-value**	**lag value**
sadness	0.019	87	9.504e-8	41
sentiment	0.003	132	0.040	137
arousal	0.111	99	0.056	134
valence	0.112	106	0.058	44
dominance	0.132	106	0.091	44
love	0.072	71	0.140	50

3.1.2 Granger causality test - Ethereum

We repeat the analysis for the Ethereum returns time series. In this case, we find significant (direct) Granger-causality between the *anger*, *sentiment*, *valence*, *arousal*, *love* and *dominance* metrics series and the Ethereum returns time series. Instead, we can conclude that *joy* and *politeness* metrics do not *Granger cause* the Ethereum returns series.

As an illustration of the process for lag selection, we show in Fig. 4 the p-values obtained for each lag value, up to the chosen τ, for the affect time series that are correlated with the Ethereum returns (direct causality).

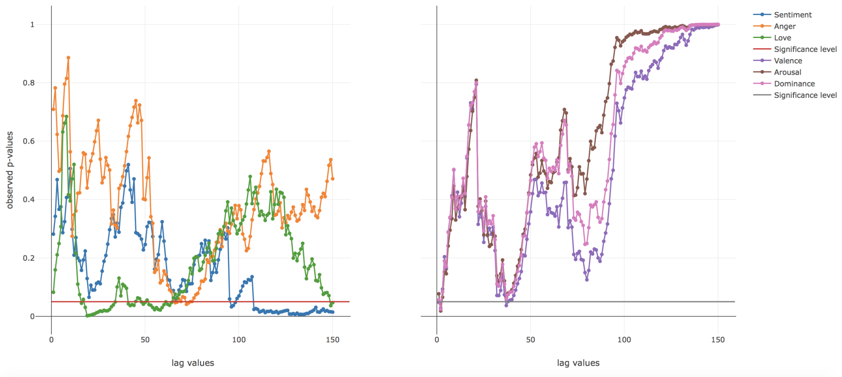


Figure 4: **Ethereum direct causality - p-values as a function of lags.** Left: p-values for the direct causality tests between sentiment, anger, love affect time series and returns. Right: p-values for the direct causality tests between valence, arousal and dominance affect time series and returns.

The *reverse Granger-causality test* results highlight, instead, that *joy*, *valence*, *arousal*, *love* and *dominance* affect metric influence the returns of Ethereum.

As for the Bitcoin analysis, we select as time lag for the Granger causality, the value associated with the minimal significant p-value. In Table 5 we provide the minimal p-values and the associated time lags for *all* affect metrics for direct and reverse causality.

As for the Bitcoin case, we compute the AIC values associated with each time lag, selecting the models with minimal AIC.

As an example, we show here the analysis for the *love* metric. In Figure 5 we report the AIC values as a function of the lag parameter. We notice that the lowest values of AIC (corresponding to minimal information loss) are recorded in correspondence with the time lag values ($\sim$ 19) associated with the lowest p-value. This analysis is, therefore, consistent with the Granger test results. Similar conclusions can be drawn for other affect time series "Granger-causing" the Ethereum returns. AIC values as a function of the time lags can be found for *all* affect metrics Granger-causing the Ethereum returns in Sec. 2 of the Supplementary Material. We have also checked of the AIC values of the models to ascertain that the Granger test was not capturing spurious effects. AIC values as a function of the time lags can be found for *all* affect metrics Granger-causing the Bitcoin returns in Sec. 2 of the Supplementary Material. As for Bitcoin, we also estimate the time lag using the BIC criterion and results are provided in the Supplementary Material (Sec. 2).

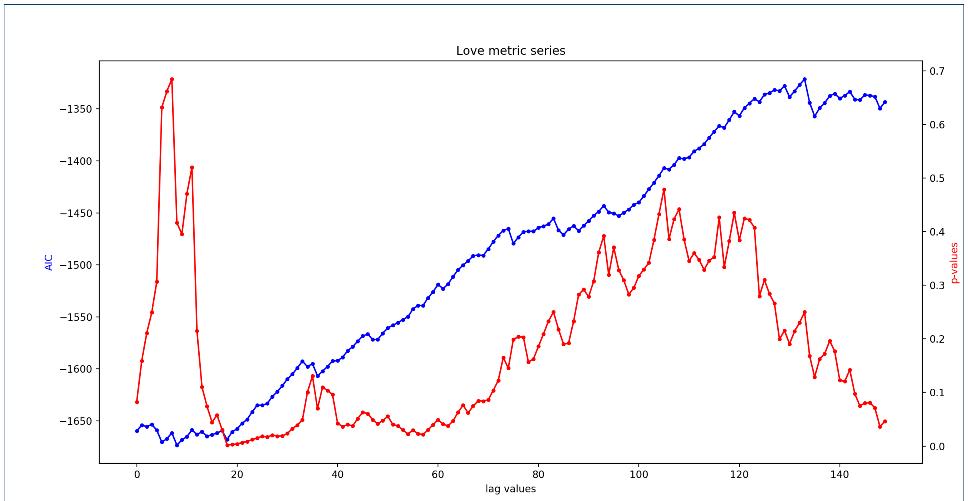


Figure 5: **Ethereum - Love AIC and p-values.** AIC values (see Eq. (6)) (blue line) and *p-values* (red lines) as a function of the number of lags.

The final set of time series showing a robust Granger (direct and/or reverse) causality with Ethereum returns according to both the p-value and the AIC tests is reported in Table 5. As before, we summarise the results of the Granger test, including time lags and the associated p-values.

To summarise, in this case, we have a *unidirectional Granger-causality* from *anger*, *sentiment*, *valence*, *arousal*, *love* and *dominance* series to Ethereum returns and a reverse *unidirectional causality* from Ethereum returns to *joy*, *love*, *valence*, *arousal* and *dominance* metric series.

3.1.3 General remarks for the Bitcoin and Ethereum analysis

In general, for both cryptocurrencies, the observed p-values are well below the chosen 5% significance level. In particular, the p-values obtained for the *sentiment* metric is even below the 1% significance level, in both the Bitcoin and Ethereum analysis.

Table 5: **Ethereum Granger Causality tests.** Minimum observed p-values and corresponding lag values for different affect time series. We highlight in green the cases showing significant (direct or reverse) Granger-causality.

	Granger Causality Test		*Reverse Granger Causality Test*	
Series	**p-value**	**lag value**	**p-value**	**lag value**
joy	0.207	4	0.011	1
love	0.002	19	0.039	113
anger	0.041	72	0.290	16
valence	0.028	2	0.025	1
arousal	0.018	2	0.040	1
sentiment	0.006	134	0.248	1
dominance	0.024	2	0.022	1
politeness	0.056	17	0.053	56

In terms of time lag, the test highlights that the Bitcoin returns time series seems to be affected by sadness metrics and developers sentiment only after a period of the order of $3-5$ months (see Tables 4, 5). Similar considerations can be made in the case of Ethereum for the *Anger* and *Sentiment* affect time series. *Dominance*, *Arousal*, *Valence* and *Love* metrics series appears, instead, to have short-term effects on the Ethereum returns time series.

We could speculate that the short-term and long-term nature of the effects of affect metrics on returns is related to the nature of cryptocurrencies itself. For instance, on the Ethereum platforms, developers can issue multiple tokens with different features and often the developers themselves are those advertising the tokens and making transactions to increase their values. As highlighted in a recent research[9] a dominant fraction of the transactions on the Ethereum blockchain appears to be handled by token teams giving new tokens for free (airdrops) to Ethereum users, therefore possibly impacting the total valuation of the platform.

In the Bitcoin case, the long-term effect of changes in developers' affect metrics may be correlated with the market efficiency. Indeed, in [10, 12] the authors show that the Bitcoin market is not efficient, i.e. that all information is not instantly incorporated into prices, hence the large time lag of the causality.

Finally, disagreements among developers of a platform may signal and lead to a fork event, which in turn generates price movements as shown in [55]. From the onset of a disagreement within the community to the actual fork attempt there is generally a significant time lag, possibly of weeks or months, compatible with our results.

Regarding the reverse causality, in the Bitcoin case we notice rather high lag values (as for the direct causality, i.e. affect metrics → Bitcoin returns), hence the Bitcoin community does not react immediately to price news. In Ethereum, instead, price movements impact the community with a time lag of 1-day. We could speculate that this effect is once again related to the different uses of the two blockchain platforms (e.g. multiple tokens issued on the Ethereum blockchain). In a related study on topic analysis of tech forums on Reddit [22], authors also find that topics on "fundamental

[9] `https://research.tokenanalyst.io/classifying-ethereum-users-using-blockchain-data/`

cryptocurrency value" are very frequent in the Ethereum community threads and are correlated with increase in prices.

3.2 Is the affect of Bitcoin and Ethereum communities able to improve the error on the prediction of returns?

As we discussed in the previous analysis, the decisions taken by the community of developers may have a non-negligible impact on the crypto-market. In this section, we further investigate the predictive power of the affect time series over the cryptocurrency returns. In particular, we use a deep learning algorithm to predict the cryptocurrency returns in two scenarios, (i) using only the cryptocurrency returns as a feature or (ii) incrementally adding the affect metrics to determine whether the additional affect metrics features yield an improvement in the prediction of the *Root Mean Square Error* (RMSE). By prediction of the RMSE we mean the average squared error of the correct estimation of the daily returns compared with the actual returns. The details of the algorithm we used were described previously in Sec. 2.4.

The results obtained for the RMSE of the predictions (measured at the end of the test phase, i.e. after 50 training epochs) are summarised in Table 6 and 7 for Bitcoin and Ethereum respectively. We compute the RMSE value by varying the number of features used in the algorithm. We consider as features the affect time series that showed direct Granger causality with the Bitcoin returns (see Table 4). For the Bitcoin analysis (Table 6), the 1-feature case corresponds to including only the time series of Bitcoin returns, while the 3-feature case includes the return time series together with the sadness and sentiment time series. We proceed in a similar way for the Ethereum case, where we incrementally include affect time series to the prediction model for the returns (considering the affect metrics that showed causality with the returns, summarised in Table 5).

Features	RMSE
1	0.129
2	0.032
3	0.013

Table 6: **Bitcoin prediction errors.** Root Mean Square Error (RMSE) of predictions considering (1) only Bitcoin returns and then sequentially adding sentiment(2) and sadness (3) time series as features.

Interestingly, we find that including the affect time series in models (based on LSTM neural networks) for the prediction of cryptocurrency returns yield a decrease in the RMSE. This result holds true for the prediction of the time series of both the Bitcoin and Ethereum returns. Indeed, in both Table 6, 7, we can see that when adding all the affects metrics, the RMSE of the predictions is significantly improved, from 0.129 to 0.013 (90% of improvement) for Bitcoin and from 0.178 to 0.048 (73% of improvement) for Ethereum.

We compared the distributions of the RMSEs for the 1-feature model (including only cryptocurrency returns) and the final model with all the features. The distributions of RMSEs include the RMSE values for each one of the 50 training epochs for

Features	RMSE
1	0.178
2	0.082
3	0.113
4	0.114
5	0.114
6	0.114
7	0.048

Table 7: **Ethereum prediction errors.** Root Mean Square Error (RMSE) of predictions considering Ethereum returns (1), then adding sequentially sentiment (2), anger (3), arousal (4), valence (5), dominance (6), love (7) time series.

the two models (including 1-feature only or all affect metrics respectively). For this comparison we used the *Wilcoxon Rank-Sum* test, a nonparametric test that does not assume specific characteristics of the distributions, e.g. normality, compared to equivalent tests (e.g. the Welch test) [56]. We find that the two distributions are statistically different with a $p - value$ of 0.0002 for Bitcoin (effect size of 0.56) and 0.00001 for Ethereum (effect size 0.48).

To summarise, we show that (i) by aggregating *all* the features (i.e. all affect metrics) we obtain – for both Bitcoin and Ethereum – a significant increase in predictive power than when considering them separately. Moreover, (ii) we provide examples of cases where also the partial aggregation (using only some of the affect metrics that Granger-cause the returns, e.g. considering Ethereum returns and the anger time series) is better than inputting only the time series of returns for the prediction task. These examples are non-exhaustive of all possible combinations of affect time series and returns as input of the neural network, but serve as illustrations that a decrease in prediction error can be induced by the addition of the affect metric.

4 Threats To Validity

Threats to external validity concern the generalisation of our results. In this study, we analysed comments from GitHub for Bitcoin and Ethereum open source projects. Our results cannot be representative of *all* other cryptocurrencies and this could, indeed, affect the generality of the study. Replication of this work on other open source cryptocurrency-related projects is needed to confirm our findings. Additionally, the politeness tool can be subject to bias due to the domain used to train the machine learning classifier.

Threats to internal validity concern confounding factors that can influence the obtained results. Based on empirical evidence, we assume a relationship between the emotional state of developers and what they write in issue reports [57]. Since the main goal of developers' communication is the sharing of information, the consequence of removing or camouflaging emotions may make comments less meaningful and cause misunderstandings. This work is focused on sentences written by developers for developers. To illustrate the influence of these comments, it is important to understand the language used by developers. We believe that all the tools used for measuring the affect metrics are valid in the software development domain. The comments used in this study were collected over an extended period from developers unaware of being monitored, therefore, we are confident that the emotions, sentiment, politeness and VAD metrics we analysed are genuine ones.

Threats to construct validity focus on how accurately the observations describe the phenomena of interest. The detection of emotions from issue reports presents difficulties due to vagueness and subjectivity. Emotions, sentiment and politeness measures are approximated and cannot perfectly identify the precise context, given the challenges of natural language and subtle phenomena like sarcasm.

5 Conclusions

Blockchain development processes have deep foundations within the community, with the community itself being the "heart and brain" of all critical decisions around the improvements and changes on the platforms. Investors and crypto-market players look at the development activities and read the technical reports of the developers to try to predict the success of the platforms they are betting on. There is, indeed, a connection between the development activities and the valuation of cryptocurrencies. In this paper, we uncovered this connection using quantitative approaches based on sentiment, politeness, emotions and VAD analysis of Github comments of two major blockchain projects, Ethereum and Bitcoin. According to our investigation affect time series do carry predictive power over the prices of cryptocurrencies. Indeed, this pioneering analysis will be extended in the near future to include other major cryptocurrencies and token development projects (e.g. ERC20 Ethereum-based tokens, ZCash or Monero) to confirm the presence of similar correlation patterns and impact of affect metrics on prices. When - in the darkness of their own rooms - blockchain developers lash out at or "wow" colleagues on Github, they might not even suspect that such simple actions could lead – months later and miles away – other people to make or lose money.

Declarations

Competing interests

The authors declare they have no competing interests.

Availability of data and materials

Data and codes used in this study are publicly available and can be found here
https://bitbucket.org/giuseppedestefanis/the-butterfly-affect/src/master/.

Funding

SB, GD and MO acknowledge funding by UCL Centre for Blockchain Technologies as part of the *1st Internal Call for Project Proposals on Distributed Ledger Technologies*.

Authors' Contributions

Silvia Bartolucci, Giuseppe Destefanis, Marco Ortu, Nicola Uras conceived and designed the experiments, performed the experiments, collected the data, analysed the data, wrote the paper, performed the computation work, reviewed drafts of the paper.
Michele Marchesi and Roberto Tonelli reviewed drafts of the paper.

Author details

[1]Imperial College Business School, London, UK. [2]Brunel University London, UK. [3]DIEE, University of Cagliari, It. [4]DMI, University of Cagliari, It. [5]UCL Centre for Blockchain Technologies, UK.

References

1. Phillips, R.C., Gorse, D.: Cryptocurrency price drivers: Wavelet coherence analysis revisited. PloS one **13**(4), 0195200 (2018)
2. Ciaian, P., Rajcaniova, M., Kancs, D.A.: The economics of bitcoin price formation. Applied Economics **48**(19), 1799–1815 (2016)
3. Cong, L.W., Ye, L., Neng, W.: Tokenomics: Dynamic adoption and valuation. Becker Friedman Institute for Research in Economics Working Paper (2018-49) (2018)
4. Bartolucci, K.A. S.: A model of the optimal selection of crypto assets. ArXiv preprint (arXiv:1906.09632) (2019)
5. Alessandretti, L., ElBahrawy, A., Aiello, L.M., Baronchelli, A.: Anticipating cryptocurrency prices using machine learning. Complexity (2018)

6. Jing-Zhi, H., William, H., Jun, N.: Predicting bitcoin returns using high-dimensional technical indicators. The Journal of Finance and Data Science (2018). doi:10.1016/j.jfds.2018.10.001
7. Lahmiri, B.S. S.: Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons and Fractals **118**, 35–40 (2019)
8. Drozdz, S., Gabarowski, R., Minati, L., Oswiecimka, P., Watorek, M.: Bitcoin market route to maturity? evidence from return fluctuations, temporal correlations and multiscaling effects. Chaos: An Interdisciplinary Journal of Nonlinear Science **28**(7), 071101 (2018). doi:10.1063/1.5036517
9. Drozdz, S., Minati, L., Oswikecimka, P., Stanuszek, M., Watorek, M.: Signatures of crypto-currency market decoupling from the forex. arXiv preprint arXiv:1906.07834 (2019)
10. Urquhart, A.: The inefficiency of bitcoin. Economics Letters **148**, 80–82 (2016)
11. Katsiampa, P.: Volatility estimation for bitcoin: A comparison of garch models. Economics Letters **158**, 3–6 (2017)
12. Lahmiri, S., Bekiros, S., Salvi, A.: Long-range memory, distributional variation and randomness of bitcoin volatility. Chaos, Solitons & Fractals **107**, 43–48 (2018)
13. Conrad, C., Custovic, A., Ghysels, E.: Long-and short-term cryptocurrency volatility components: A garch-midas analysis. Journal of Risk and Financial Management **11**(2), 23 (2018)
14. Walther, T., Klein, T., Bouri, E.: Exogenous drivers of bitcoin and cryptocurrency volatility–a mixed data sampling approach to forecasting. University of St. Gallen, School of Finance Research Paper (2018/19) (2019)
15. Bouri, E., Lau, C.K.M., Lucey, B., Roubaud, D.: Trading volume and the predictability of return and volatility in the cryptocurrency market. Finance Research Letters **29**, 340–346 (2019)
16. Kim, Y.B., Kim, J.G., Kim, W., Im, J.H., Kim, T.H., Kang, S.J., Kim, C.H.: Predicting fluctuations in cryptocurrency transactions based on user comments and replies. PLOS ONE **11**(8), 1–17 (2016). doi:10.1371/journal.pone.0161197
17. Li, T.R., Chamrajnagar, A.S., Fong, X.R., Rizik, N.R., Fu, F.: Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. Frontiers in Physics **7**, 98 (2019). doi:10.3389/fphy.2019.00098
18. Aste, T.: Cryptocurrency market structure: connecting emotions and economics. special issue of digital finance on cryptocurrencies. Digital Finance (2018)
19. Keskin, Z., Aste, T.: Information-theoretic measures for non-linear causality detection: application to social media sentiment and cryptocurrency prices. arXiv:1906.05740 (2019)
20. Chen, C.Y.-H., Hafner, C.M.: Sentiment-induced bubbles in the cryptocurrency market. Journal of Risk and Financial Management **12**(2), 53 (2019)
21. Yelowitz, A., Wilson, M.: Characteristics of bitcoin users: an analysis of google search data. Applied Economics Letters **22**(13), 1030–1036 (2015)
22. Phillips, R.C., Gorse, D.: Mutual-excitation of cryptocurrency market returns and social media topics. In: Proceedings of the 4th International Conference on Frontiers of Educational Technologies., pp. 80–86 (2018). ACM
23. Graziotin, D., Wang, X., Abrahamsson, P.: Understanding the affect of developers: theoretical background and guidelines for psychoempirical software engineering. In: Proceedings of the 7th International Workshop on Social Software Engineering - SSE 2015, pp. 25–32. ACM Press, New York, New York, USA (2015). doi:10.1145/2804381.2804386. http://dl.acm.org/citation.cfm?doid=2804381.2804386
24. Destefanis, G., Ortu, M., Counsell, S., Swift, S., Marchesi, M., Tonelli, R.: Software development: do good manners matter? PeerJ Computer Science **2**, 73 (2016)
25. Murgia, A., Ortu, M., Tourani, P., Adams, B., Demeyer, S.: An exploratory qualitative and quantitative analysis of emotions in issue report comments of open source systems. Empirical Software Engineering, 1–44 (2017)
26. Graziotin, D., Wang, X., Abrahamsson, P.: Happy software developers solve problems better: psychological measurements in empirical software engineering. PeerJ **2**, 289 (2014)
27. Khan, I.A., Brinkman, W.-P., Hierons, R.M.: Do moods affect programmers' debug performance? Cognition, Technology & Work **13**(4), 245–258 (2011)
28. Ong, B., Lee, T.M., Li, G., Chuen, D.L.K.: Evaluating the potential of alternative cryptocurrencies. In: Handbook of Digital Currency, pp. 81–135. Elsevier, ??? (2015). doi:https://www.sciencedirect.com/science/article/pii/B9780128021170000059
29. Islam, Z.M.F. M. R.: Towards understanding and exploiting developers' emotional variations in software engineering. In: 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), pp. 185–192 (2016). doi:10.1109/SERA.2016.7516145
30. de Albornoz, J.C., Plaza, L., Gervás, P.: Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis. In: LREC, pp. 3562–3567 (2012)
31. Mantyla, M., Adams, B., Destefanis, G., Graziotin, D., Ortu, M.: Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity? In: Proceedings of the 13th International Conference on Mining Software Repositories, pp. 247–258 (2016). ACM
32. Russell, J.A.: Emotion, core affect, and psychological construction. Cognition & Emotion **23**(7), 1259–1283 (2009). doi:10.1080/02699930902809375
33. Graziotin, D., Wang, X., Abrahamsson, P.: How do you feel, developer? an explanatory theory of the impact of affects on programming performance. PeerJ Computer Science **1**, 18 (2015)
34. Ortu, M., Hall, T., Marchesi, M., Tonelli, R., Bowes, D., Destefanis, G.: Mining communication patterns in software development: A github analysis. In: Proceedings of the 14th International Conference on Predictive Models and Data Analytics in Software Engineering, pp. 70–79 (2018). ACM
35. Murgia, A., Tourani, P., Adams, B., Ortu, M.: Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In: Proceedings of the 11th Working Conference on Mining Software Repositories, pp. 262–271 (2014). ACM
36. Murgia, A., Ortu, M., Tourani, P., Adams, B., Demeyer, S.: An exploratory qualitative and quantitative analysis of emotions in issue report comments of open source systems. Empirical Software Engineering **23**(1), 521–564

(2018)
37. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods **45**(4), 1191–1207 (2013). doi:10.3758/s13428-012-0314-x
38. Danescu-Niculescu-Mizil, S.M.J.D.L.J.P.C. C.: A computational approach to politeness with application to social factors. In: Proceedings of ACL (2013)
39. Calefato, F., Lanubile, F., Maiorano, F., Novielli, N.: Sentiment polarity detection for software development. Empirical Software Engineering **23**(3), 1352–1382 (2018)
40. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society **37**, 424–438 (1969)
41. Thornton, D.L., Batten, D.S.: Lag-length selection and tests of Granger causality between money and income. Journal of Money, credit and Banking **17**(2), 164–178 (1985)
42. Liew, V.K.-S.: Which lag length selection criteria should we employ? Economics bulletin **3**(33), 1–9 (2004)
43. Balcilar, M., Bouri, E., Gupta, R., Roubaud, D.: Can volume predict bitcoin returns and volatility? a quantiles-based approach. Economic Modelling **64**, 74–81 (2017)
44. Schwarz, G., *et al.*: Estimating the dimension of a model. The annals of statistics **6**(2), 461–464 (1978)
45. Gonzalo, J., Pitarakis, J.-Y.: Lag length estimation in large dimensional systems. Journal of Time Series Analysis **23**(4), 401–423 (2002)
46. Jones, J.D.: A comparison of lag–length selection techniques in tests of Granger causality between money growth and inflation: evidence for the US, 1959–86. Applied Economics **21**(6), 809–822 (1989)
47. Seabold, S., Perktold, J.: Statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference (2010)
48. Fahrmeir, L., Kneib, T., Lang, S., Marx, B.: Regression, Springer, (2007)
49. Banks, H.T., Joyner, M.L.: Aic under the framework of least squares estimation. Applied Mathematics Letters **74**, 33–45 (2017)
50. Roman, J., Jameel, A.: Backpropagation and recurrent neural networks in financial analysis of multiple stock market returns. In: Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences, vol. 2, pp. 454–460 (1996). IEEE
51. Dase, R.K., Pawar, D.D.: Application of artificial neural network for stock market predictions: A review of literature. International Journal of Machine Intelligence **2**(2), 14–17 (2010)
52. McNally, S., Roche, J., Caton, S.: Predicting the price of bitcoin using machine learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), pp. 339–343 (2018). IEEE
53. Chen, Z., Li, C., Sun, W.: Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics **365**, 112395 (2020)
54. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
55. Chaim, P., Laurini, M.P.: Volatility and return jumps in bitcoin. Economics Letters **173**, 158–163 (2018)
56. Fay, M.P., Proschan, M.A.: Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules. Statistics surveys **4**, 1 (2010)
57. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval **2**(1-2), 1–135 (2008)