Chaos

Using networks and partial differential equations to forecast bitcoin price movement

Cite as: Chaos **30**, 073127 (2020); https://doi.org/10.1063/5.0002759
Submitted: 28 January 2020 . Accepted: 02 July 2020 . Published Online: 17 July 2020

Yufang Wang, and Haiyan Wang

Chaos ARTICLE scitation.org/journal/cha

Using networks and partial differential equations to forecast bitcoin price movement

Cite as: Chaos **30**, 073127 (2020); doi: 10.1063/5.0002759
Submitted: 28 January 2020 · Accepted: 2 July 2020 ·
Published Online: 17 July 2020

View Online Export Citation CrossMark

Yufang Wang[1] and Haiyan Wang[2,a)]

AFFILIATIONS

[1]School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China
[2]School of Mathematical and Natural Sciences, Arizona State University, Tempe, Arizona 85069, USA

[a)]**Author to whom correspondence should be addressed:** haiyan.wang@asu.edu

ABSTRACT

Over the past decade, the blockchain technology and its bitcoin cryptocurrency have received considerable attention. Bitcoin has experienced significant price swings in daily and long-term valuations. In this paper, we propose a partial differential equation (PDE) model on the bitcoin transaction network for forecasting the bitcoin price movement. Through analysis of bitcoin subgraphs or chainlets, the PDE model captures the influence of transaction patterns on the bitcoin price over time and combines the effect of all chainlet clusters. In addition, Google Trends index is incorporated to the PDE model to reflect the effect of the bitcoin market sentiment. The experiment results demonstrate that the PDE model is capable of forecasting the bitcoin price movement. The paper is the first attempt to apply a PDE model to the bitcoin transaction network for forecasting.

Bitcoin is the world's leading cryptocurrency and its value often undergoes large swings over short periods. Given the high volatility and the difference from traditional currencies, the price of bitcoin is extremely difficult to forecast. In this paper, we aim to propose a partial differential equation (PDE) model to forecast the bitcoin price movement. New contributions of this paper are as follows: (i) it is the first attempt to establish a PDE model on the bitcoin transaction network for the bitcoin price dynamics; (ii) we aggregate similar transaction patterns in the bitcoin transaction network into clusters and embed the clusters into a Euclidean space; (iii) within the framework of the PDE forecasting model, we integrate Google Trends and chainlets to account for the effect of chainlet clusters; and (iv) the forecasting accuracy of the bitcoin price direction is above 50% in each 3 months of 2007.

I. INTRODUCTION

Bitcoin is currently the world's leading cryptocurrency, and the blockchain is the technology that underpins it. The concept of bitcoin was first suggested in 2008 by Satoshi Nakamoto, and it became fully operational in January 2009.[1] By May 2018, the market capitalization of bitcoin had arrived at nearly \$115 billion. In contrast to the traditional financial asset, whose records of everyday monetary transactions are considered highly sensitive and are kept private, bitcoin has no financial intermediaries, and a complete list of its transactions is publicly available in a public ledger. This publicly distributed ledger creates opportunities for people to observe all the financial interactions on the blockchain network and analyze how the assets circulate in time.

The bitcoin value (i.e., the price of a bitcoin) often undergoes large swings over short periods. For instance, at the beginning of 2013, the price started at nearly \$13 per bitcoin and then rocketed to \$230 on 9 April, yielding almost 1700% profits in less than 4 months; in 2017, the value of a single bitcoin increased 2000%, going from \$863 on January 9, 2017 to a high of \$17 550 on December 11, 2017.

Given the high volatility and the difference from traditional currencies, the price of bitcoin is extremely difficult to forecast. A few studies have been conducted on forecasting or estimation for bitcoin prices. Regression models are the mostly used methods for bitcoin price forecasting by considering some potential price-affecting factors. For instance, Ciaian *et al.*[2] forecast the bitcoin price with a linear regression model by considering some factors in marco-finance and attractiveness for investors. Jang and Lee[3] introduce blockchain information (such as hash rate and block generation rate) to increase the forecasting accuracy by a Bayesian neural network. Researchers also apply machine learning techniques to make bitcoin price forecasting.[4–7] Atsalakis *et al.*[7] use "a hybrid neuro-fuzzy controller, namely, PATSOS, to forecast the direction in the change of the daily price of bitcoin." Cretarola and Figà-Talamanca[8] propose an ordinary difference equation to describe the behavior of the bitcoin price by considering investors' attention for bitcoin. As the publicly available ledger of the bitcoin network can be represented by a directed graph, some works exist to make bitcoin forecasting through network analysis.[5,9] Kurbucz[5] forecasts the price of bitcoin by the most frequent edges of its transaction network. This study shows that the utility of global graph features can be used to forecast the bitcoin price.

It is generally accepted that the bitcoin price is significantly affected by attention or sentiment about the bitcoin system itself.[10–12] The frequency of searches for the term "bitcoin" in Google Trends has proved to be a good measure of interest.[8,10,13] Cretarola and Figà-Talamanca[8] regard the Google Trends index as a proxy for the attention measure and, thus, propose a bivariate model in continuous time to describe the behavior of the bitcoin price. Kristoufek[10] demonstrates quantitatively that not only are the Google Trends index and the prices connected but a pronounced asymmetry also exists between the effect of increased interest in the currency while it is above or below its trend value.

Recently, Akcora *et al.*[9] studied the influence of local topological structures on the bitcoin price dynamics. They combined "chainlets" of the bitcoin transaction networks with statistical models to forecast the bitcoin price. Essentially, chainlets are special forms of network motifs or subgraphs in the address-transaction bitcoin graph. Chainlets describe transactions occurring in a blockchain and each chainlet represents a trading decision or transaction pattern. In Ref. 9, Akcora *et al.* found that certain types of chainlets exhibit an important role in bitcoin price forecasting. It has been found that the bitcoin price is mainly and strongly linked with transaction activities.[12,14] Analysis of motifs of the bitcoin transaction networks has been found to be an indispensable tool to unveil hidden mechanisms of bitcoin networks for the bitcoin price dynamics.

In this paper, we aim to propose a partial differential equation (PDE) model to forecast the bitcoin price movement. As indicated in Ref. 9, the predictive utility of different types of transactions for the bitcoin price dynamics may be different. As a result, we apply spectral analysis to aggregate bitcoin transaction subgraphs (chainlets) into clusters with similar types of transaction patterns. We embed the clusters of chainlets into a Euclidean space and develop a PDE model to incorporate the bitcoin market sentiment with the Google Trends index. There exist many dynamic models describing information diffusion in the online social network, for instance, of Refs. 15–17. Furthermore, the framework of PDE models developed by the authors for information diffusion in online social networks in Refs. 18–20 is adapted to the bitcoin transaction network for characterizing the effect of the chainlet clusters. The PDE model enables us to describe the influence of the clusters on the price over time. To assess the forecasting ability of our model, hit ratio and relative accuracy are applied to measure the forecasting accuracy in the perspectives of the bitcoin price direction and bitcoin price. Numerical results demonstrate that the PDE model is capable of forecasting the bitcoin price movement. The paper is the first attempt to apply a PDE model on the bitcoin transaction network for forecasting the bitcoin price movement.

The remainder of the paper is organized as follows: some basic concepts are introduced in Sec. II. The PDE-based forecasting model is proposed in Sec. III. The forecasting process is described in Sec. IV. Section V concludes the paper with a discussion.

II. BITCOIN AND CHAINLET

A blockchain is a distributed ledger that records transactions in blocks without requiring a trusted central authority. Each block contains a set of transactions and it has a hash link to its previous block, thus creating a chain of chronologically ordered blocks. When transactions happen at the same time, they will be recorded in the same block.

Bitcoin addresses are used for receiving bitcoins. A transaction represents the flow of bitcoins from input address to output addresses over time. A transaction is multi-input and multi-output, which means that a transaction may have more than one input address and more than one output address. Users take part in the bitcoin economy through addresses, and a user can have two or more addresses at the same time.

The transaction–address graph[9,21] is a directed graph, which is vital for knowing the bitcoin flowing state. In this graph, the set $\{Address, Transaction\}$ consists of vertices and edges, in which an edge represents the bitcoin's transfer between and address node and a transaction node as in Fig. 1. As transactions are the only means to manage bitcoins, so bitcoins can be divided or aggregated only by being spent. For instance, a transaction involves multiple addresses and every user can have different addresses, so the user can use a transaction to split, merge, or move bitcoins between its own addresses. Therefore, each transaction with its input and output addresses represents a decision, encoded by a subgraph in the transaction–address graph.

The subgraph, composed of a transaction with its input and output addresses, is called *1-chainlet*, or simply may be called *chainlet*, which is a novel graph data model, proposed in 2018[9] for studying the bitcoin price dynamics. A chainlet with x inputs and y outputs is often noted as $C_{x\longrightarrow y}$. Chainlets have distinct shapes

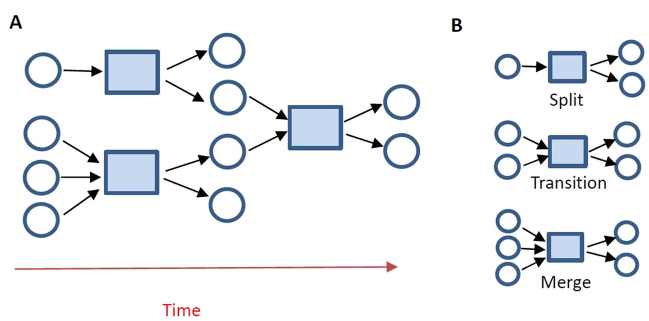

FIG. 1. (a) A transaction–address graph; (b) split ($C_{1\longrightarrow 2}$), transition ($C_{2\longrightarrow 2}$), and merge ($C_{3\longrightarrow 2}$) chainlets. The three types, merge, transition, and split, are determined according to the relative number of input addresses and output addresses and correspond to the state that the former is greater than, equal to, or less than the latter, respectively. Addresses and transactions are shown with circles and rectangles, respectively. An arrow indicates a transfer of bitcoins.

that reflect their role in the network. They are the building blocks in blockchain analysis; they represent different transaction patterns and reflect different decisions. In this paper, we analyze bitcoin chainlets and forecast the bitcoin price dynamics.

III. PDE MODEL BASED ON CHAINLETS AND GOOGLE TRENDS

In this section, we establish a PDE model for bitcoin price forecasting. This PDE model is based on chainlets and the Google Trends index for the bitcoin price. Different types of chainlet clusters represent different transaction patterns and transaction decisions; thus, they provide different predictive utility for the bitcoin price. The PDE model proposed below captures the influence of the chainlet clusters and combines the effects of all the clusters for forecasting the bitcoin price.

A. Chainlets clustering

In this paper, we use chainlets as the building blocks for bitcoin price forecasting, and each chainlet represents a type of immutable decision. To make a better forecasting, we use spectral clustering to aggregate chainlets into clusters with similar type of transactions.

1. Chainlet network

A transaction with inputs and outputs composes a chainlet. Just as in Ref. 9, we denote $C_{x\longrightarrow y}$ as a chainlet if this chainlet has x inputs and y outputs. Different chainlets have different values of x and y (x and y are positive integers). Though the bitcoin protocol limits the number of input and output addresses for a transaction, the number of inputs and outputs can still reach thousands. As a result, millions of different chainlets occur (e.g., $C_{2000\longrightarrow 20}$ and $C_{2000\longrightarrow 120}$). However, an analysis of the entire bitcoin history shows that 97.57% of the chainlets have fewer than 20 inputs and outputs.[9] This means that in bitcoin blocks, a sufficiently large number of chainlets satisfy $1 \le x < 20, 1 \le y < 20$.

Therefore, we build a weighted graph $G(V, E)$ with 400 nodes, where each node in V is a kind of chainlet or chainlet set and the graph node set $\{\mathbb{C}_{x\longrightarrow y}, 1 \le x \le 20, 1 \le y \le 20\}$ is defined as

$$\mathbb{C}_{x\longrightarrow y} = \begin{cases} C_{x\longrightarrow y}, & \text{if } x < 20 \quad \text{and} \quad y < 20, \\ \{C_{x\longrightarrow j}, 20 \le j < +\infty\}, & \text{if } x < 20 \quad \text{and} \quad y = 20, \\ \{C_{i\longrightarrow y}, 20 \le i < +\infty\}, & \text{if } x = 20 \quad \text{and} \quad y < 20, \\ \{C_{i\longrightarrow j}, 20 \le i < +\infty, 20 \le j < +\infty\}, & \text{if } x = 20 \quad \text{and} \quad y = 20. \end{cases}$$

An edge between i and j in E and its weight are determined by the Pearson correlation coefficient of relative historical daily transaction volumes, with a correlation threshold θ.

2. Chainlets clustering

Different transactions (transactions with distinct inputs and outputs) contribute differently for the bitcoin price formation. Four hundred distinct chainlets exist in chainlet network G and each chainlet represents a type of transaction form. However, intuitively, it is not necessary to distinguish all chainlets. For example, the transaction patterns $\mathbb{C}_{3\longrightarrow 1}$ and $\mathbb{C}_{4\longrightarrow 1}$ have no meaningful difference. Therefore, we cluster chainlets by spectral clustering, which divides the graph by using the eigenvectors of the Laplacian matrix.[22] Thus, each chainlet cluster represents certain similar types of transaction pattern, which may have different predictive utilities for the bitcoin price. Our PDE model proposed in Subsection III B is capable of capturing the influence of these chainlet clusters and integrating the effects of all clusters for bitcoin price forecasting.

In this paper, we apply the daily transaction volumes of the 400 kinds of chainlets from December 1, 2016 to December 30, 2016 (denoted here it as Data-set 1) to build the chainlet network G with a Pearson correlation threshold $\theta = 0.6$, which is an empirical value to ensure that most of chainlets in the chainlet network are connected and not isolated. In fact, we can use data in other time periods to build network as well, as long as the time interval of Data-set 1 is earlier than the period for which we want to make forecasting for the bitcoin price. For example, if we want to forecast the bitcoin price in February, 2017, we can first build the network based on the data in January, 2017. We obtain 10 chainlet clusters by applying spectral clustering method to the above chainlet network G. Our forecasting period is from January 1, 2017 to December 31, 2017 (denoted it as Data-set 2). Figure 2 shows the average transaction volumes of all the chainlet clusters for this period.

The bitcoin price is mainly and strongly linked with transaction activities, especially transaction volumes.[12,14] Figure 2 shows that the chainlet clusters we obtain have different levels of transaction volumes.

B. Modeling bitcoin price with PDE

In this section, we develop a PDE model to model the influence of chainlet clusters with different transaction patterns and further combine their effects on the bitcoin price, for the purpose of forecasting the bitcoin price.

We apply the authors' framework of PDE models for information diffusion in online social networks[18–20] to the bitcoin transaction network for characterizing the influence of the chainlet clusters on the bitcoin price.

To apply a PDE model to describe the interactions of the chainlet clusters, one must embed the chainlet clusters $U_1, U_2, \ldots, U_n$ (n is the number of chainlet clusters) into a Euclidean space and arrange

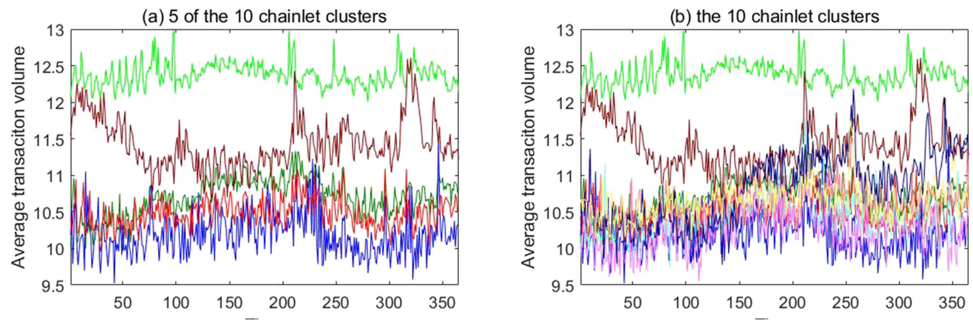

FIG. 2. (a) Average transaction volumes for five of the ten chainlet clusters; (b) the ten chainlet clusters obtained by the spectral clustering method. Each color represents a cluster. The time period is from January 1, 2017 to December 31, 2017.

them in a meaningful order. In this paper, the chainlet clusters are mapped onto a line such that connected clusters stay as close as possible, as in Fig. 3. Specifically, we now treat each chainlet cluster as a node in a new graph G_{new}, where the strength of an edge is the summation of all the weights between the two clusters. The Fiedler vector (the eigenvector corresponding to the second smallest eigenvalue) of the Laplacian matrix of the new graph G_{new} can map these chainlet clusters onto a line, showing the order as $U_{i_1}, U_{i_2}, \ldots, U_{i_n}$.[18] On this line, connected nodes stay as close as possible, ensuring that the continuous model can capture the influence of these chainlet clusters.

Having embedded chainlet clusters to the Euclidean space, let $u(x,t)$ represent the effect of chainlet cluster x on the bitcoin price. The formulation of the spatiotemporal model for the bitcoin network follows the balance law: the rate at which a given quantity changes in a given domain must equal the rate at which it flows across its boundary plus the rate at which it is created, or destroyed, within the domain. The PDE model can be conceptually divided into two processes: an internal process within each chainlet cluster and external process among chainlet clusters. Similar derivation for the PDE model has been used in our previous work for PDE models for information diffusion in online social networks in Refs. 18–20. Our proposed PDE-based model is

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(d(x)\frac{\partial u(x,t)}{\partial x}\right) + r(t)u(x,t)h(x), \tag{1}$$

where $r(t)u(x,t)h(x)$ describes the rate of change of the bitcoin price within the cluster x; $r(t)$ represents the rate of change with respect to t; $h(x)$ describes the spatial heterogeneity of different chainlet clusters or transaction patterns. $\frac{\partial}{\partial x}\left(d(x)\frac{\partial u(x,t)}{\partial x}\right)$ reflects the rate of change of the bitcoin price among chainlet clusters, $d(x)$ describes the interaction of chainlet clusters. Because $u(x,t)$ represent the influence of chainlet cluster x on the bitcoin price, we have

$$\text{Forecasted bitcoin price} = \int u(x,t)\,dx. \tag{2}$$

Investor sentiment about bitcoin, which can be represented by the Google Trends index, is a key factor for determining the bitcoin price. It is plausible to assume that

$$u(x,t) \equiv b_0 m(x,t) + \alpha(x), \tag{3}$$

where $m(x,t)$ is the predictive utility of the Google Trends index of chianlet cluster x; b_0 is a scale factor to the bitcoin price. $\alpha(x)$ describes the heterogeneity of different chainlet clusters on the bitcoin price. In this way, chainlet clusters with larger trading volumes do not necessarily have more influence on the bitcoin price.

We derive a PDE model with respect to $m(x,t)$. Note both $h(x)$ and $\alpha(x)$ represent the spatial heterogeneity of different chainlet clusters, and, therefore, we can assume that $h(x) = k\alpha(x)$ with a constant k. Substituting (3) for (1) and (2) and assuming $d(x) \equiv d > 0$, d is a constant, the forecasting model of the bitcoin price is

$$\begin{cases} \frac{\partial m(x,t)}{\partial t} = d\frac{\partial^2 m}{\partial x^2} + k\alpha(x)r(t)\Big(m(x,t) + \frac{1}{b_0}\alpha(x)\Big) + \frac{d}{b_0}\alpha''(x), \\ m(x,1) = \phi(x), L_1 < x < L_2, \\ \frac{\partial m}{\partial x}(L_1,t) = \frac{\partial m}{\partial x}(L_2,t) = 0, t > 1, \\ \text{Forecasted bitcoin price at time } t = \int_{L_1}^{L_2}\Big(b_0 m(x,t) + \alpha(x)\Big)dx, \end{cases} \tag{4}$$

where

- $\alpha(x)$ satisfies $\alpha(x_i) = \alpha_i, i = 1, 2, \ldots, n$, where α_i describes the heterogeneity of the cluster at location x_i and n is the number of clusters. We construct $\alpha(x)$ by cubic spline interpolation[23] with the condition of $\frac{\partial\alpha}{\partial x}(L_1) = \frac{\partial\alpha}{\partial x}(L_2) = 0$. Therefore, the second derivative $\alpha''(x)$ exists and is continuous; $r(t)$ satisfies exponential

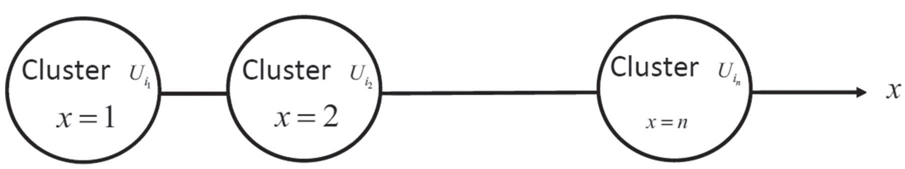

FIG. 3. Embedding of chainlet clusters into the x-axis.

decay with time. In this work, $r(t)$ is assumed the form of $r(t) = b_1 + e^{-(t-b_2)}$.

- The Neumann boundary condition $\frac{\partial m}{\partial x}(L_1, t) = \frac{\partial m}{\partial x}(L_2, t) = 0,\ t > 1$ is applied, and it has been assumed no flux of information flow across the boundaries at $x = L_1, L_2$; initial function $m(x, 1) = \phi(x)$ describes the influence of every chainlet cluster, and it is constructed from the historical data by cubic spline interpolation.[23]
- Parameters $d, b_0, b_1, b_2, \alpha_i, i = 1, 2, \ldots, n$ are determined by the known historical data of $m(x_i, t_j)$ with the best fitting program as the detailed illustration in Sec. IV B.
- The historical forecasting utility of the Google Trends index on the bitcoin price

$$m(x_i, t_j), \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, N$$

can be obtained by computing

$$m(x_i, t_j) = (\text{Google Trends index on "bitcoin" at time } t_j) * P_0(x_i, t_j),$$

where

$$P_0(x_i, t_j) = \frac{\text{Bitcoin transaction volume of chianlet cluster } x_i \text{ at time } t_j}{\text{Total bitcoin transaction volume of all chianlet clusters at } t_j}.$$

IV. FORECASTING OF BITCOIN PRICE

A. Data

As mentioned in the introduction, the bitcoin price is related to the transaction volumes and the investors' attention measured by Google.[11] The daily bitcoin transaction volumes of all transactions and the daily bitcoin price (here it refers to the intraday open price, denominated in USD in our work) are downloaded from https://github.com/cakcora/coinworks. The time period is from January 1, 2017 to December 31, 2017. These datasets are extracted from the original bitcoin data, which are all publicly available at the Bitcoin Core page (https://bitcoin.org/en/download). The Google Trends index on "bitcoin" captures the attention of retail/uniformed investors.[13] These data can be obtained from https://trends.google.it/trends/?geo=IT with a keyword "bitcoin."

We chose the year of 2017 for validating our model mainly because there was some huge price fluctuation in 2017. Specifically, the value of a single bitcoin increased 2000%, going from \$863 on January 9, 2017 to a high of \$17 550 on December 11, 2017. Thus, our model is actually also validated in time windows with a structural break in data.

B. Forecasting process

To forecast the bitcoin price at time t_{N+1}, the forecasting process consists in determination of parameters in (4) using known historical data $\{m(x_i, t_j), i = 1, 2, \ldots, n; j = 1, 2, \ldots, N\}$ and solving the PDE model (4) to make one-step forecasting for $m(x, t_{N+1}), x \in [L_1, L_2]$. Thus, the forecasted bitcoin price at time t_{N+1} is given by

$$Price(t_{N+1}) = \int_{L_1}^{L_2} \Big(b_0 m(x, t_{N+1}) + \alpha(x)\Big) dx. \tag{5}$$

Specifically, we combine a tensor train (TT) global optimization approach[24] and the Nelder–Mead simplex local optimization method[25] to train the PDE parameters. After each determination of the model parameters, we apply t"pdepe" function in MATLAB to compute the PDE for one-step forward in time dimension numerically.

In this work, we obtain ten chainlet clusters, therefore, $n = 10, L_1 = x_1 = 1, L_2 = x_{10} = 10$. Further, we use one-step sliding window of length 3 to perform forecasting process. Namely, we use historical data of $m(x_i, t_j)$ during each 3 days to forecast the bitcoin price of the 4th day. Specifically, data of days 1–3, 2–4, 3–5, ... are used as the training set for forecasting the bitcoin price of days 4, 5, 6, ..., successively. We make a forecasting for the whole year of 2017 and the forecasting results cover 362 days from January 4, 2017 to November 31, 2017.

C. Forecasting results

In general, due to the difficulty, forecasting the "bitcoin price direction" (or the "direction of the bitcoin price movement") is preferred than forecasting the "bitcoin price."[7] In this present study, we measure the predictive ability of our model with hit ratio and relative accuracy (RA).

A hit ratio measures the model ability for forecasting the "bitcoin price direction," defined as

$$\text{Hit ratio} = \frac{1}{n}\sum_{1}^{n} D_i, \quad i = 1, 2, \ldots, n,$$

where $\sum_1^n D_i$ denotes the number of correct forecasts of the bitcoin price direction

$$D_i = \begin{cases} 1, & (P_{real}(i+1) - P_{real}(i))\left(P_{forecast}(i+1) - P_{real}(i)\right) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where n is the number of tests. In this study, we compute the hit ratios every 3 months (shown as Table I). The number of 3 months usually corresponds to a 3-month investment-trading horizon, and this period has been suggested as benchmark period for evaluating a trading strategy.[26,27]

The relative accuracy (RA) for the ith trading day, defined as

$$\text{RA}(i) = 1 - \frac{|P_{real}(i) - P_{forecast}(i)|}{P_{real}(i)},$$

TABLE I. With sliding window of length 3, the statistics of the forecasting results for the bitcoin price direction from January 4, 2017 to December 31, 2017.

		Hit ratio (%)
2017	January 4–March 31	52.22
	April 1–June 30	56.32
	July 1–September 30	54.35
	October 1–December 31	54.35

is used to measure the forecasting accuracy of the bitcoin price, where $P_{real}(i)$ is the real bitcoin price for the ith trading day and $P_{forecast}(i)$ is the forecasted bitcoin price for the ith trading day.

In our forecasting time period, though the observed real price range is large from \$775.98 to \$19 498.68, the forecasted values of our proposed forecasting model well capture the trend of the real bitcoin price, as in Fig. 4(a). As expected, the forecasting performance of the proposed model deteriorates as the observed real data series is skyrocketing, as in November and December of 2017 in Fig. 4(a). However, through statistical analysis of the forecasting results, in the last 3 months of 2017, the hit ratio is still above 50%; of the 61 days in the last 2 months of 2017, there exist 25 days and 37 days whose relative accuracies were more than 0.8 and 0.7, respectively.

Specifically, Table I shows that the overall hit ratios (i.e., the percentage of accurate forecasts for the direction of price movement) are all higher than the threshold of 50% by chance (i.e., flip of a coin). In Ref. 6, it illustrates that by flipping a fair and random coin, an investor has a 50% chance to be right or wrong about the market direction. Although the hit ratios we obtained are not too much larger than 50%, the entire story will be different if we are interested in a sequence of events. Table II also shows that the overall average forecasting accuracy of 362 consecutive days in 2017 can reach 0.82. All the forecasting results are shown in Fig. 4(b). Of the 362 consecutive days in 2017, 82% ($296/362 \approx 82\%$), 65% ($237/362 \approx 65\%$), and 37% ($134/362 \approx 37\%$) of the days achieve an accuracy of above 0.7, 0.8, and 0.9, respectively. These statistics for the forecasting results are summarized in Table II.

To further calibrate our model, we also apply sliding windows of lengths 6 and 9 to perform one-step forward out-of-sample forecasting, in the present analysis. The statistics of days for the forecasting results are shown in Tables III and IV (using sliding windows of length 6) and Tables V and VI (using sliding windows of length 9). With the increasing of the window length from 3, 6, to 9, both hit ratios and RA have decreasing trend. This may be due to high fluctuation of the daily bitcoin price. Nevertheless, the forecasting results for the three windows are acceptable. Here, we also emphasize that the lengths of the sliding windows considered by us are for illustrative purpose. One can easily change the length of the window according to different types of data. However, empirical results demonstrate that data of only several days are needed for an acceptable forecasting.

V. DISCUSSION

In this paper, a PDE model is developed for forecasting the bitcoin price dynamics based on daily bitcoin transaction volumes and Google Trends index. Experiment results demonstrate that the

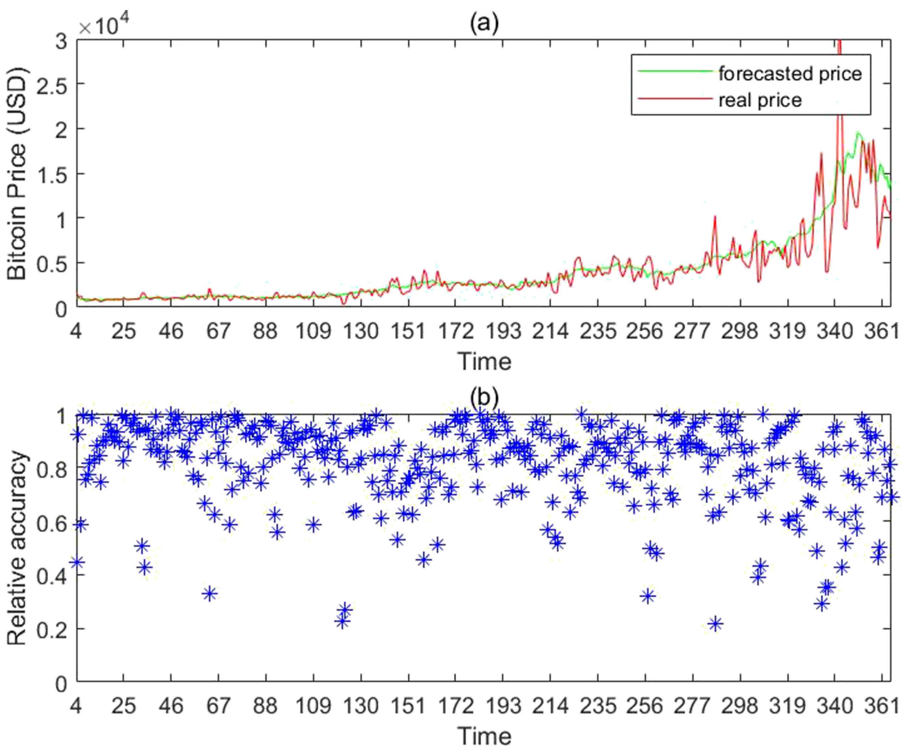

FIG. 4. Forecasted results of the bitcoin daily price from January 4, 2017 to December 31, 2017. (a) Forecasted values vs the real observations. (b) The relative accuracy for bitcoin price forecasting. Here, the relative accuracy (RA) is the conventional definition as $RA = 1 - \frac{|P_{real} - P_{forecast}|}{P_{real}}$, where P_{real} is the real bitcoin price at every data collection time point and $P_{forecast}$ is the forecasted bitcoin price through (4).

TABLE II. With sliding window of length 3, the statistics of days for the forecasting results of the bitcoin price from January 4, 2017 to December 31, 2017. 362 and 134 mean that during the forecasting period of 362 days, there are 134 days whose relative accuracy of the forecasting is above 0.9.

Total days	362	Average relative accuracy (RA)=0.82
Days of RA > 0.9	134	(Days of RA > 0.9)/(Total days) = 37%
Days of RA > 0.8	237	(Days of RA > 0.8)/(Total days) = 65%
Days of RA > 0.7	296	(Days of RA > 0.7)/(Total days) = 82%

TABLE III. With sliding window of length 6, the statistics of the forecasting results for the bitcoin price direction from January 7, 2017 to December 31, 2017.

		Hit ratio (%)
2017	January 7–March 31	51.19
	April 1–June 30	52.75
	July 1–September 30	51.09
	October 1–December 31	52.17

TABLE IV. With sliding window of length 6, the statistics of days for the forecasting results of the bitcoin price from January 7, 2017 to December 31, 2017. 359 and 108 mean that during the forecasting period of 359 days, there are 108 days whose relative accuracy of the forecasting is above 0.9.

Total days	359	Average relative accuracy (RA) = 0.77
Days of RA > 0.9	108	(Days of RA > 0.9)/(Total days) = 30%
Days of RA > 0.8	194	(Days of RA > 0.8)/(Total days) = 54%
Days of RA > 0.7	267	(Days of RA > 0.7)/(Total days) = 74%

TABLE V. With sliding window of length 9, the statistics of the forecasting results for bitcoin price direction from January 10, 2017 to December 31, 2017.

		Hit ratio (%)
2017	January 10–March 31	51.85
	April 1–June 30	51.65
	July 1–September 30	51.09
	October 1–December 31	51.09

TABLE VI. With sliding window of length 9, the statistics of days for the forecasting results of the bitcoin price from January 10, 2017 to December 31, 2017. 356 and 89 mean that during the forecasting period of 356 days, there are 89 days whose relative accuracy of the forecasting is above 0.9.

Total days	356	Average relative accuracy (RA) = 0.74
Days of RA > 0.9	89	(Days of RA > 0.9)/(Total days) = 25%
Days of RA > 0.8	161	(Days of RA > 0.8)/(Total days) = 45%
Days of RA > 0.7	199	(Days of RA > 0.7)/(Total days) = 56%

forecasting accuracy of the bitcoin price direction (measured by hit ratio) is above 50% in each 3 months of 2007; the average forecasting accuracy (measured by relative accuracy) of our model is 0.82 for 362 consecutive days in 2017. Because of different datasets, it may be not comparable with other works. Nevertheless, Kurbucz[5] achieves an accuracy of approximately 60.05% during daily price movement classifications between November 25, 2016 and February 5, 2018. Jiang and Lee[3] obtain an acceptable forecasting accuracy through selecting different relevant features of blockchain information and comparing the Bayesian neural network with benchmark models on modeling.

Our work differs from the previous forecasting models with chainlets. Akcora *et al.*[9] introduce "chainlet" on blockchain for bitcoin price forecasting, but they focus only on the effects of certain types of chainlets on the bitcoin price. Our proposed PDE model emphasizes the combined effects from all the different types of chainlet clusters. Furthermore, the continuous PDE model describes the influence of these chainlet clusters over time.

In addition, our PDE model differs from our previous PDE models on social networks for forecasting information diffusion,[18] air pollution of 189 cities in China[19] and influenza prevalence.[20] In this paper, our PDE model is developed to capture the combined effect of chainlets from the bitcoin transaction network and their influence on the bitcoin price. In particular, the Google Trends index is incorporated in the model to reflect the effect of market sentimental. Unlike the base linear or logistic models in Refs. 18–20, the PDE model in this paper has additional terms to describe the spatial heterogeneity of chainlet clusters. As a result, chainlet clusters with larger trading volumes do not necessarily have more influence on the bitcoin price. In fact, the forecasting of the bitcoin price is based on the combined influence of all chainlet clusters.

ACKNOWLEDGMENTS

Y.W. was partially supported by the Humanities and Social Science Fund of Ministry of Education of China (18YJCZH184), the Natural Science Foundation of Tianjin City (Tianjin Natural Science Foundation) (19JCQNJC14800), and the National Social Science Fund Youth Project (19CGL002).

APPENDIX: MATHEMATICAL COMPUTATIONS

1. Spectral clustering

We assume $G = (V, E)$ be an undirected weighted graph with vertex set $V = \{v_1, \ldots, v_n\}$. The weighted adjacency matrix of the graph is matrix $W = (w_{ij})_{i,j=1,\ldots,n}$, where $w_{ij} \geq 0$ is the non-negative weight between vertices v_i and v_j. If $w_{ij} = 0$, this means vertices v_i and v_j are not connected. The degree matrix D of Graph G is defined as the diagonal matrix with the degrees $d_1, \ldots, d_n$ on the diagonal, where $d_i = \sum_{j=1}^{n} w_{ij}$.

Given a set of vertices $V = \{v_1, \ldots, v_n\}$ that we want to cluster into k subsets, the algorithm of spectral clustering is as follows:

1. Form the weighted adjacency matrix W and compute the Laplacian matrix $L = D^{-1/2}(D - W)D^{-1/2}$.
2. Find $x_1, x_2 \ldots, x_k$, the k largest eigenvectors of L (chosen to be orthogonal to each other in the case of repeated eigenvalues)

and form the matrix $X = [x_1, x_2, \ldots, x_k] \in \mathbb{R}^{n\times k}$ by stacking the eigenvectors in columns.

3. Form the matrix Y from X by renormalizing each of X's rows to have unit length (i.e., $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)^{1/2}$).
4. Treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters via K-means (this method is used in the present work) or any other algorithm (that attempts to minimize distortion).
5. Finally, assign the original vertex v_i to cluster j is and only if row i of the matrix Y was assigned to cluster j.

The detailed algorithm of spectral clustering is provided in Refs. 18 and 22.

2. Cubic spline interpolation

Cubic spine interpolation, corresponding "spline" function in MATLAB software, is a method fitting piecewise polynomials to data points. This interpolation method ensures that the second derivatives of these polynomials are continuous at join-points. Suppose that we have a set of $n+1$ points (which do not have to be evenly spaced),

$$(x_i, y_i), \quad i = 0, 1, \ldots, n.$$

Cubic spline fits a sets of 3rd-degree polynomials, between each pair of points, from x_i to x_{i+1}. The points at which the splines join are called knots. Each spline (each 3rd-degree polynomial) must join with its neighboring cubic polynomials at the knots where they join with the same slope (to ensure 1st-derivatives of the interpolation function are continuous) and curvature (to ensure end-derivatives of the interpolation function are continuous).

Specifically, we write the equation for a cubic polynomial, $g_i(x)$, in the ith interval, between points (x_i, y_i), (x_{i+1}, y_{i+1}). It has the form $g_i(x) = a_i(x-x_i)^3 + b_i(x-x_i)^2 + c_i(x-x_i) + d_i$. Thus, the cubic spline function we want is of the form

$$g(x) = g_i(x), \quad x \in [x_i, x_{i+1}], \quad \text{for} \quad i = 1, 2, \ldots, n-1$$

and meets these conditions,

$$\begin{cases} g_i(x_i) & = y_i, \quad i = 0, 1, \ldots, n-1 \quad g_{n-1}(x_n) = y_n, \\ g_i(x_{i+1}) & = g_{i+1}(x_{i+1}), \quad i = 0, 1, \ldots, n-2, \\ g_i^{'}(x_{i+1}) & = g_{i+1}^{'}(x_{i+1}), \quad i = 0, 1, \ldots, n-2, \\ g_i^{''}(x_{i+1}) & = g_{i+1}^{''}(x_{i+1}), \quad i = 0, 1, \ldots, n-2. \end{cases}$$

For the detailed algorithm of cubic spine interpolation refer to Ref. 23.

3. Numerical solution of parabolic partial differential equation (PDE)

Many problems in science and engineering are modeled by special cases of the parabolic equation. The general form of an one-dimensional case is

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(d(x,t)\frac{\partial u(x,t)}{\partial x}\right) + f(x,u,t),$$

where x and t are the spatial and temporal variables, respectively.[23] Many numerical approaches are often used to solve the parabolic equations. For instance, the explicit method of finite-difference, at point x_i and time t_j, $\frac{\partial u}{\partial t} \approx \frac{u_i^{j+1} - u_i^j}{\Delta t}$ is used for approximating time derivative, and $\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{(\Delta x)^2}$ is applied for approximating the derivative with respect to x, where the subscripts and superscripts indicate the location and time, separately. Often to obtain high numerical solution of PDE, multiple numerical methods will be applied for the numerical experiments. In the present work, we apply the "pdepe" function in MATLAB software to solve our proposed PDE.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Nakamoto *et al.*, see www.bitcoin.org for "Bitcoin: A peer-to-peer electronic cash system."

[2] P. Ciaian, M. Rajcaniova, and D. Kancs, "The economics of bitcoin price formation," Appl. Econ. **48**(19), 1799–1815 (2016).

[3] H. Jang and J. Lee, "An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information," IEEE Access **6**, 5427–5437 (2017).

[4] S. Velankar, S. Valecha, and S. Maji, "Bitcoin price prediction using machine learning," in *2018 20th International Conference on Advanced Communication Technology (ICACT)* (IEEE, 2018), pp. 144–147.

[5] M. T. Kurbucz, "Predicting the price of bitcoin by the most frequent edges of its transaction network," Econ. Lett. **184**, 108655 (2019).

[6] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," J. Comput. Appl. Math. **365**, 112395 (2020).

[7] G. S. Atsalakis, I. G. Atsalaki, F. Pasiouras, and C. Zopounidis, "Bitcoin price forecasting with neuro-fuzzy techniques," Eur. J. Oper. Res. **276**, 770–780 (2019).

[8] A. Cretarola and G. Figà-Talamanca, "Modeling bitcoin price and bubbles," in *Cryptocurrencies* (IntechOpen, 2018).

[9] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu, "Forecasting bitcoin price with graph chainlets," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer, Cham, 2018), Vol. 10939, pp. 765–776.

[10] L. Kristoufek, "Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era," Sci. Rep. **3**, 3415 (2013).

[11] J. Bukovina and M. Martiček *et al.*, "Sentiment and bitcoin volatility," Technical Report, Mendel University in Brno, Faculty of Business and Economics, 2016.

[12] L. Kristoufek, "What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis," PLoS One **10**(4), e0123923 (2015).

[13] J. Engelberg and P. Gao, "In search of attention," J. Finance **66**(5), 1461–1499 (2011).

[14] D. Koutmos, "Bitcoin returns and transaction activity," Econ. Lett. **167**, 81–85 (2018).

[15] Q. Shao, C. Xia, L. Wang, and H. Li, "A new propagation model coupling the offline and online social networks," Nonlinear Dyn. **98**, 2171–2183 (2019).

[16] C. Xia, Z. Wang, C. Zheng, Q. Guo, Y. Shi, M. Dehmer, and Z. Chen, "A new coupled disease-awareness spreading model with mass media on multiplex networks," Inf. Sci. **471**, 185–200 (2019).

[17] Z. Wang, Q. Guo, S. Sun, and C. Xia, "The impact of awareness diffusion on SIR-like epidemics in multiplex networks," Appl. Math. Comput. **349**, 134–147 (2019).

[18] H. Wang, F. Wang, and X. Kuai, *Modeling Information Diffusion in Online Social Networks with Partial Differential Equations* (Springer, 2020).

[19] Y. Wang, H. Wang, S. Chang, and A. Avram, "Prediction of daily PM_2.5 concentration in China using partial differential equations," PLoS One **13**(6), e0197666 (2018).

[20]F. Wang, H. Wang, K. Xu, R. Raymond, J. Chon, S. Fuller, and A. Debruyn, "Regional level influenza study with geo-tagged Twitter data," J. Med. Syst. **40**(8), 189 (2016).
[21]C. G. Akcora, M. F. Dixon, Y. R. Gel, and M. Kantarcioglu, "Bitcoin risk modeling with blockchain graphs," Econ. Lett. **173**, 138–142 (2018).
[22]U. Von Luxburg, "A tutorial on spectral clustering," Stat. Comput. **17**(4), 395–416 (2007).
[23]C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis* (Addison-Wesley, 1994).
[24]I. V. Oseledets, "Tensor-train decomposition," SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011).
[25]J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," SIAM J. Optim. **9**(1), 112–147 (1998).
[26]G. S. Atsalakis and P. V. Kimon, "Forecasting stock market short-term trends using a neuro-fuzzy based methodology," Expert Syst. Appl. **36**(7), 10696–10707 (2016).
[27]G. S. Atsalakis, D. Frantzis, and C. Zopounidis, "Commodities' price trend forecasting by a neuro-fuzzy controller," Energy Syst. **7**, 73–102 (2016).