

“Summary of Forecasting Bitcoin Price with Graph Chainlets[2]”

Cuneyt G. Akcora, Asim Kumer Dey, Yulia R. Gel and Murat Kantarcioglu
(Blockchain Data Analytics, University of Manitoba, Fall 2020)

1. INTRODUCTION

Over the last several years, bitcoin’s [19] has received a tremendous amount of attention from both the academic community and outside academia because of its wide array of applications[23] and many other reasons. The most exciting feature that bitcoin employs is its distributed ledger. With this feature employed in bitcoin, all the users that intend to do transactions can use this ledger to verify whether a transaction is authentic or not. With this distributed ledger and using graph analysis, we might be able to explore some exciting and novel applications.

This paper aims to use the bitcoin graph for addressing a very important issue in bitcoin, which is the price forecast. The idea of using the bitcoin graph had been used for the analysis of the stock market. There have been some studies on the role of small subgraphs or motifs about how financial networks might function and provide early warnings stability indicators[11]. So this paper’s main contributions would be the introduction of chainlets motifs to analyze the environment of bitcoin price and come up with approaches for comparing them and see which one holds more promise. With these contributions, the authors prove that the tool they have introduced has an outstanding performance regarding price forecast.

Chainlets: A bitcoin graph is comprised of three components: addresses, transactions, and blocks. So let’s begin defining the bitcoin graph: Bitcoin network graph is a directed graph (V, E, B) with V representing the set of vertices of the graph, and $E \in V * V$ represents the set of edges in the graph. $B = \{Address, Transaction\}$ defines the type of vertex in the graph, and adjacent nodes in this graph should not have the same type.

k-chainlets: The authors introduce a concept called k-chainlets in order to assess the local higher order topological structure of the Bitcoin graph. So assume that the graph $G_k = (V_k, E_k, B)$ (k-chainlet) is a subgraph of G and that it has k nodes that are $\{transaction\}$. We denote the number of frequencies of G_k in G by $f_g(G_k)$ by k-chainlet signature. From now on, we refer to 1-chainlets as chainlets.

For a particular chainlet, if it has a inputs and b outputs we show this by $C_{a \rightarrow b}$. If the number of inputs are higher than the number of outputs we call these **merge** chainlets $x_{a \rightarrow b}$ such that $a > b$. There are two other chainlets which are **transition** and **split** such that $a = b$ and $a < b$. From now on we call these chainlets **Aggregate Chainlets**. Figure 3 represents the percentage of aggregate chainlets over time.

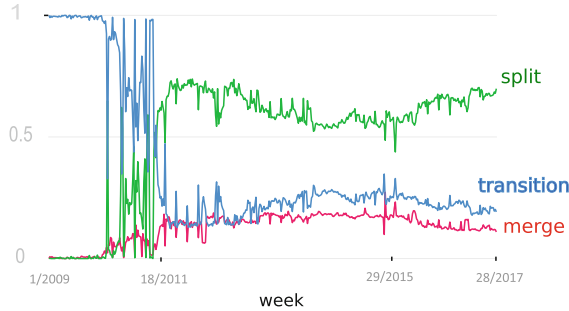


Fig. 3. Percentage of aggregate chainlets in weeks. Splits constitute around 60% of all transactions.

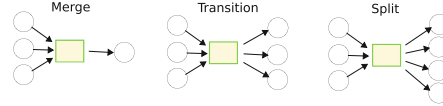


Fig. 2. Merge ($C_{3 \rightarrow 1}$), Transition ($C_{3 \rightarrow 3}$) and Split ($C_{3 \rightarrow 4}$) chainlets for 3 inputs.

The authors use a matrix presentation in order to model the bitcoin graph with chainlets over time. After the construction of the bitcoin graph, chainlet counts obtained from this graph would be saved in a matrix. To select the matrix's right dimension, the authors performed an analysis of the complete bitcoin history and decided to choose 20 for the matrix dimension. Furthermore, following that, the authors took daily and weekly snapshots with 3.284 and 4000 matrices, and they show every 400 chainlets as a vector. Moreover, the authors clustered chainlets over chainlet vectors, which you could see that in figure 4. White cells are chainlets that form a cluster of itself.

There are three important rules with regard to this graph. The first rule states that it is possible to merge multiple transactions and spend it in a single transaction. And the second rule implies that the mappings related to input-output addressed are not explicitly recorded in this graph. The third rule is that the coins from multiple transactions can be spent separately, whereas the coins from one transaction must be spent in a single transaction. In the network graph, transactions are ordered by blocks so that every transaction illustrates an irrevocable decision, which is encoded as a subgraph. Furthermore, the authors prefer to use these subgraphs over edges and vertices for their approach. They call these subgraphs chainlets. This was due to two reasons.

At first, the authors begin explaining how chainlets could have a decisive role in predicting using a concept called Granger causality. Granger causality is a statistical concept of causality that is based on prediction. According to Granger causality, if a signal X_1 "Granger-causes" (or "G-causes") a signal X_2 , then past values of X_1 should contain information that helps predict X_2 above and beyond the information contained in past values of X_2 alone. Its mathematical formulation is based on linear regression modeling of stochastic processes (Granger 1969). More complex extensions to nonlinear cases exist. However, these extensions are often more difficult to apply in practice[1]. After that, they demonstrate how chainlets could have a decisive role in bitcoin price prediction.

Findings and future directions: We mentioned that the authors used an economic concept called Granger causality to see how chainlets could be used to forecast bitcoin price. By using this concept and the fact that the authors have access to the time series of chainlets, they want to investigate how the structure of the bitcoin graph would affect the bitcoin prices in the future, and they want to know that if chainlets themselves could hold some information regarding the bitcoin prices in the future. In the end, they were able to prove that chainlets G-cause the

bitcoin price, or, i.e., they have a predictive effect on prices.

The authors state that in order to evaluate the performance of chainlets, they could use any model they want but mention that some models like Random Forest(RF) models perform better than BJ models in terms of prediction. So the authors conducted large-scale experiments with six RF models. The baseline model is comprised of lagged(past period) values of bitcoin price.

In figure 5, you could observe the comparison of percent decrease for different models with prediction horizons, which are not constant and begin from 1 to 30. The authors observed that for 1-step ahead forecast, chainlets do not provide helpful predictions. However, on the other hand, from 3 steps ahead and more chainlets began to have an influential role in terms of prediction. In addition to that, the authors also observed that some chainlets seem to have more utility in price prediction compared to others.

In figure 6, you could see a comparison with the baseline model and three other models(1,2 and 5). For 1-step ahead forecast, all of them seem to perform similarly, and as you would expect, the performance of all models decreased when the step ahead forecast increases too much($h \rightarrow \infty$). Nevertheless, even in this case, models 1,2, and 5 had a slightly better performance in comparison to the baseline model.

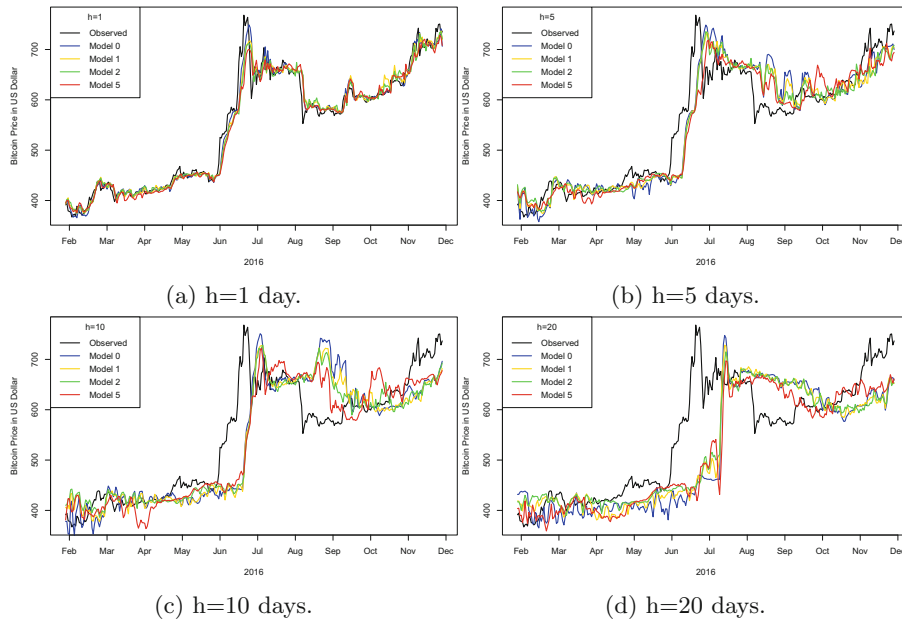


Fig. 6. [Color Online]. Price prediction for 2016 with 1, 5, 10 for 20 day horizons.

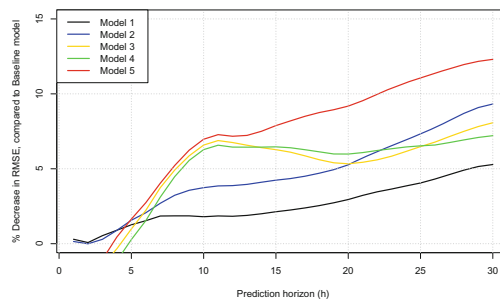


Fig. 5. % Change (decrease) in RMSE compared to the baseline model.

REFERENCES

- [1] Anil Seth (2007) Granger causality. Scholarpedia, 2(7):1667., revision #127333
- [2] Akcora, Cuneyt G., et al. "Forecasting Bitcoin price with graph chainlets." Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2018.
- [3] Ahmed, Nesreen K., et al. "Graphlet decomposition: Framework, algorithms, and applications." Knowledge and Information Systems 50.3 (2017): 689-722.
- [4] Akcora, Cuneyt Gurcan, Yulia R. Gel, and Murat Kantarcioglu. "Blockchain: A graph primer." arXiv preprint arXiv:1708.08749 (2017).
- [5] Androulaki, Elli, et al. "Evaluating user privacy in bitcoin." International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg, 2013.
- [6] Baumann, Annika, Benjamin Fabian, and Matthias Lischke. "Exploring the Bitcoin Network." WEBIST (1). 2014.
- [7] Di Battista, Giuseppe, et al. "Bitconeview: visualization of flows in the bitcoin transaction graph." 2015 IEEE Symposium on Visualization for Cyber Security (VizSec). IEEE, 2015.
- [8] Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3), 424-438 (1969)
- [9] Greaves, Alex, and Benjamin Au. "Using the bitcoin transaction graph to predict the price of bitcoin." No Data (2015).
- [10] Huang, Anna. "Similarity measures for text document clustering." Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand. Vol. 4. 2008.
- [11] Jiang, X. F., T. T. Chen, and B. Zheng. "Structure of local interactions in complex financial dynamics." Scientific reports 4 (2014): 5321.
- [12] Kane, Michael J., et al. "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks." BMC bioinformatics 15.1 (2014): 276.
- [13] Kondor, Dániel, et al. "Inferring the interplay between network structure and market effects in Bitcoin." New Journal of Physics 16.12 (2014): 125003.
- [14] Kondor, Dániel, et al. "Do the rich get richer? An empirical analysis of the Bitcoin transaction network." PloS one 9.2 (2014): e86197.
- [15] Lischke, Matthias, and Benjamin Fabian. "Analyzing the bitcoin network: The first four years." Future Internet 8.1 (2016): 7.
- [16] Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms." URL: <http://cs229.stanford.edu/proj2014/Isaac>
- [17] Milo, Ron, et al. "Network motifs: simple building blocks of complex networks." Science 298.5594 (2002): 824-827.
- [18] Möser, Malte, Rainer Böhme, and Dominic Breuker. "An inquiry into money laundering tools in the Bitcoin ecosystem." 2013 APWG eCrime researchers summit. Ieee, 2013.
- [19] Nakamoto, Satoshi, and A. Bitcoin. "A peer-to-peer electronic cash system." Bitcoin.—URL: <https://bitcoin.org/bitcoin.pdf> (2008).
- [20] Ober, Micha, Stefan Katzenbeisser, and Kay Hamacher. "Structure and anonymity of the bitcoin transaction graph." Future internet 5.2 (2013): 237-250.
- [21] Portnoff, Rebecca S., et al. "Backpage and bitcoin: Uncovering human traffickers." Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.
- [22] Sorgente, Mariano, and Cristian Cibils. "The reaction of a network: Exploring the relationship between the bitcoin network structure and the bitcoin price." No Data (2014).
- [23] Tschorsch, Florian, and Björn Scheuermann. "Bitcoin and beyond: A technical survey on decentralized digital currencies." IEEE Communications Surveys Tutorials 18.3 (2016): 2084-2123.
- [24] White, Halbert, Karim Chalak, and Xun Lu. "Linking granger causality and the pearl causal model with settable systems." NIPS Mini-Symposium on Causality in Time Series. 2011.
- [25] Yang, Steve Y., and Jinhyoung Kim. "Bitcoin market return and volatility forecasting using transaction network flow properties." 2015 IEEE Symposium Series on Computational Intelligence. IEEE, 2015.