2015 IEEE Symposium Series on Computational Intelligence

Bitcoin Market Return and Volatility Forecasting Using Transaction Network Flow Properties

Steve Y. Yang
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, New Jersey 07030
Email: steve.yang@stevens.edu

Jinhyoung Kim
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, New Jersey 07030
Email: jkim18@stevens.edu

***Abstract*—Bitcoin, as the foundation for a secure electronic payment system, has drawn broad interests from researchers in recent years. In this paper, we analyze a comprehensive Bitcoin transaction dataset and investigate the interrelationship between the flow of Bitcoin transactions and its price movement. Using network theory, we examine a few complexity measures of the Bitcoin transaction flow networks, and we model the joint dynamic relationship between these complexity measures and Bitcoin market variables such as return and volatility. We find that a particular complexity measure of the Bitcoin transaction network flow is significantly correlated with the Bitcoin market return and volatility. More specifically we document that the residual diversity or freedom of Bitcoin network flow scaled by the total system throughput can significantly improve the predictability of Bitcoin market return and volatility.**

I. Introduction

Bitcoin is a peer-to-peer decentralized crypocurrency system proposed by Satoshi Nakamoto[20]. Unlike a traditional currency, Bitcoin relies on the cryptography to secure transactions, and there is no single institution to control Bitcoin transactions. Bitcoin transactions are broadcasted by clients into a peer-to-peer system and get confirmed once they are added to the ‘block chain’, a chain of blocks containing all historical transactions since the Bitcoin’s inception. The Bitcoin ledger, the ’block chain’, is publicly visible and is stored in all Bitcoin clients to prevent the double-spending issue. This publicly available transaction information provides a unique opportunity to study the dynamic behavior of this financial asset.

The Bitcoin transaction network can be described as a network of transaction flows among all market participants. Buyers and sellers (from the peer-to-peer system as well as from the Bitcoin exchanges) of Bitcoin are the nodes of the network, and the edges represent transactions among these participants. Bitcoin transaction network therefore represents all activities within the Bitcoin market, and they are recorded in the ‘block chain’. The Bitcoin’s exchange rate is determined primarily by supply and demand of the network, and users of this virtual currency can spend Bitcoin on both virtual and real goods and services. Our hypothesis is that as the network flow of the Bitcoin transactions evolves over time, the flow of these transactions should reflect its market value and its volatility due to the information conveyed through the supply-and-demand of the system. Hence, how to measure the flow of supply-and-demand of this financial asset holds the key to explaining its market prices.

Network analysis is a natural tool to understanding complex social and economic phenomena. Like any financial market transaction networks, there is always an amount associated with the individual Bitcoin transactions at any point of time. Therefore, it is natural to consider the Bitcoin transaction networks as weighted rather than unweighted networks. Ecologists have developed a set of variables to quantify the growth and development of ecological or economic flow networks [24]; these variables have the potential to give quantitative expression to many of the qualitative observations. However the extent to which unweighted networks can be used to describe the real world network flow problems is limited. In real systems, flows have unequal sizes with sometimes extraordinary differences [26]. And hence the question becomes “What characterizes the flow among the vertices?”

In this study, we investigated two complexity measures for weighted network in order to quantify the movements of flows within a network as a time-series, and we aim to reveal the inherent interrelationships between these network complexity measures and the Bitcoin market variables. The primary measures used in this study are the numbers of flows and nodes, connectivity in flows per node, and the number of roles as measured by nodes divided by connectivity. The primary contribution of this paper is to document that there exist persistent interrelationships between the Bitcoin transaction network flows and market quality measures, and this discovery will be very valuable for quantifying risk dynamics involved in a Bitcoin based electronic payment system in the future.

The rest of this paper is organized as follows. Section II provides a literature review of the existing research on network metrics and Bitcoin market. Section III presents the Bitcoin market and the Bitcoin transaction data. It explains the mechanism of the ‘block chain’ transactions. Section IV describes the methodology to construct the Bitcoin transaction flow network and the complexity measures for a flow network. Section V introduces a prediction model of the Bitcoin market using the network complexity measures. Section VI discusses the performance of the predictive model. And section VII concludes the findings and points to future research directions.

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.248

II. Literature review

Networks are well characterized both structurally and quantitatively by graph theory, which has more than 150 years of extensive development and application ([17],[4],[4], [5], [6], and [3]). Graph theory as a branch of discrete mathematics has been brought to life to solve specific problems from different areas of science. Quantifying centrality and connectivity helps us identify portions of the network that may play interesting roles. Researchers have been proposing metrics for centrality for the past 50 years. Centrality metrics are designed to"characterize the importance of a vertex?" "Importance" can be conceived in relation to a type of flow or transfer across the network. This allows centralities to be classified by the type of flow they consider important [5]. Centralities are either Radial or Medial. Radial centrality measures count walks which start/end from a given node. The degree and eigenvalue centralities are the examples of Radial centralities, counting the number of walks of length one or length infinity [16]. Medial centralities count walks which pass through a given node. The typical example is Freedman's betweenness centrality, the number of shortest paths which pass through a given node [8].

The applications of flow networks are numerous in fields such as ecology [1], [28], economics [13], and of course, engineering ([19],[7], and [2]). It should be noted that although weighted flow networks are identical in form to weighted digraphs, the convention in the literature is that digraph weights represent costs or lengths, and thus larger weights on an edge indicate lesser significance. In flow networks, larger weights on a flow represent larger flows and thus greater importance. Kauffman has related connectivity of boolean logic networks to their stability ([10],[11]). May was one of the firsts to argue that the stability of a complex system is related to the connectance of a trophic web [15]; others disputed the form of the relationship, but generally agreed that connectivity relates to stability ([21],[9]). Ulanowicz and Zorach developed a weighted flow network measure to quantify the complexity of weighted networks [25]. They presented a consistent way to generalize the measures of vertices, flows, connectivity, and roles into weighted networks, and they argued that weighting leads in the end to a more elegant and fruitful analysis of networks in complex systems.

A number of studies have applied network analysis on the Bitcoin public ledger. Reid and Harrigan studied anonymity of the Bitcoin network and analyzed the topology of the two types of networks: the transaction network and the user network [20]. These networks were constructed from Bitcoin transactions occured between 1/3/2009 and 5/13/2012. In order to construct a user network, they proposed a data pre-processing step to identify a Bitcoin user as a node of the user network. To improve the identification accuracy in the user network, the authors used the external information where Bitcoin users disclosed their public addresses online. They showed that it is possible to map the Bitcoin public address to IP addresses and then link them to the previous transactions. From the proposed techniques, they investigated the alleged theft of Bitcoin in the user network. Ron and Shamir constructed a Bitcoin transaction network in the same way as the user network used by Reid and Harrigan [22]. The authors calculated various statistics such as distributions of addresses, the amount of incoming Bitcoin, balances of a Bitcoin public addresses, the number and size of transactions, and the most active entities. They found that the majority of the minted Bitcoin are not in use and a large number of tiny transactions exist. And they noticed that hundreds of transactions transfer more than 50,000 Bitcoin(BTC) and found that most of these big transactions are successors of the initial ones. Another interesting finding is that transaction flows reveal certain distinct behaviors such as long chains, fork merge, and binary tree-like distributions. Ober et al. studied time-varying dynamics of the network structure and the degree of anonymity [18]. Using data from Bitcoin's inception to 1/6/2013, the authors discovered that the entity sizes and the overall pattern of usage became more stationary in the last 12 to 18 months of the period in the study, which reduces the anonymity set. The authors also showed that the number of dormant coins is important to quantify anonymity. Inactive entities hold many of these dormant coins and thus further reduce the anonymity set. Kondor et al. studied the structure of Bitcoin transaction network and its time evolution [12]. They divided the lifetime of the Bitcoin system into two phases based on network properties such as degree distribution, degree correlation, and clustering. The first phase in time, called initial phase, is characterized by large fluctuation of network properties. During this phase, Bitcoin is primarily used as an experiment rather than a real currency. The second phase, called trading phase, shows more stable network properties, and Bitcoin thus draws more public attention and becomes a real currency. They also studied wealth distribution over all Bitcoin public addresses and found that the wealth distribution is quite heterogeneous during the entire lifetime but become stable in the trading phase.

III. Data

In this section, we describe the dataset used in this research. We collected the historical Bitcoin transactions and Bitcoin market prices. Both of them are available at the website www.blockchain.info.

A. Bitcoin Transaction

This study covers a broad Bitcoin transaction history, whereas previous studies only dealt with transactions occurred when Bitcoin was not actively traded on exchanges. The Bitcoin public ledger records all historical transactions in the form of a chain of blocks. We collected blocks up to the block number 336,859 which contain over 55 million transactions. Our dataset covers all Bitcoin transactions from the Bitcoin's inception until 12/31/2014. Here, a Bitcoin transaction is defined as the change in ownership of Bitcoin between the public addresses. A single transaction consists of two parts, the input and the output. An input of a Bitcoin transaction specifies the public addresses and the amount of Bitcoin stored in the associated public addresses. An output specifies the public addresses and the amount of Bitcoin transferred into the associated public addresses. A Bitcoin transaction may have multiple addresses in both the input and the output. The Fig. 1 shows a real transaction with 2 addresses in the input and 2 addresses in the output. The amount of Bitcoin is represented in Satoshi which is the smallest unit of the Bitcoin currency recorded on the block chain. One Satoshi is a one hundred millionth of a single Bitcoin i.e. 1 Bitcoin is equivalent to 10^8 Satoshi.

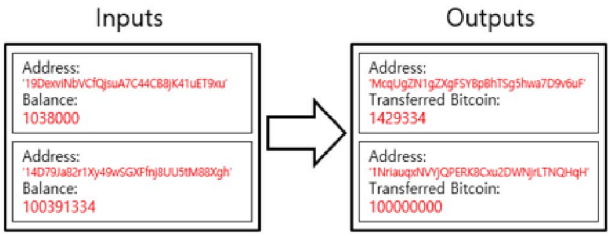

Fig. 1: Bitcoin transaction # 2642247 generated at 1/1/2012 6:22:01

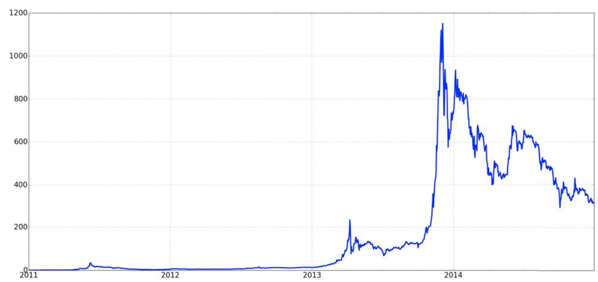

Fig. 2: Bitcoin market prices in USD from 1/1/2011 through 12/31/2014

TABLE I: Basic statistics of the daily Bitcoin market price return and volatility.

	Max	Min	Mean	S.D.	Skewness
return	0.641853	-0.478305	0.004767	0.063737	1.279526
volatility	0.192331	0.042228	0.063947	0.023509	2.138615

B. Bitcoin Market

Fig. 2 exhibits the time series of the Bitcoin market price from 1/1/2011 through 12/31/2014. The first Bitcoin exchange, Mt. Gox, was launched in July 2010 and Bitcoin was traded under 5 USD before May 2011. It started to draw public attention in 2013. It passed a price of $1000 all-time high on November 28, 2013 and fell to around $400 in April 2014. The price quickly rebounded to $600 and then steadily declined to no more than $200.

Daily Bitcoin price return is defined as the log difference between index values of two consecutive days. And we estimate the Bitcoin price volatility using a GARCH(1,1) model. In Fig. 3, we plot the time series of daily Bitcoin price return and volatility. The basic statistics for the time series are summarized in TABLE I.

IV. Methodology

A. Bitcoin Network

We construct a Bitcoin transaction network for each day during the sample period by following the methodology used by Reid and Harrigan [20]. In Reid and Harrigan, a node of network is defined as a group of Bitcoin public addresses which are assumed to be owned by the same entity. Every Bitcoin address of a node appearing in the input part of a Bitcoin transaction and is involved with one address in the same node through a Bitcoin transaction. This method is justifiable because public addresses involved in the input part of the same transaction should be owned by the same entity before they can be used in a single Bitcoin transaction. To identify the ownership of public addresses, we apply the Union-Find algorithm to a set of 59,151,261 public addresses and identify over 5 million entities. And we then assign a user identification number to each entity. An edge of the Bitcoin network indicates that there is a transaction between two entities. Edge has a directionality from the entity on the input side to the entity on the outputs side. Since there could be multiple entities on the output side of a transaction, multiple edges could occur as a result of a single transaction. If the entity in the outputs is the same as the entity in the inputs, we then will ignore this self-loop. A weight on an edge indicates the amount of transferred Bitcoin through an associated transaction. Since there could be multiple transactions from one entity to another entity during a day, the weight on an edge is the sum of weights from the multiple transactions. Fig. 4 shows a processed Bitcoin transaction with a user ID number and part of a Bitcoin transaction network. Before the transaction #2642247, the user #1777953210 had 10.1429334 BTC in his public addresses. This user sends 10 BTC out of 10.1429334 he owes to the user #38667927. And the remaining change of 0.1429334 BTC is kept by user #17779532.

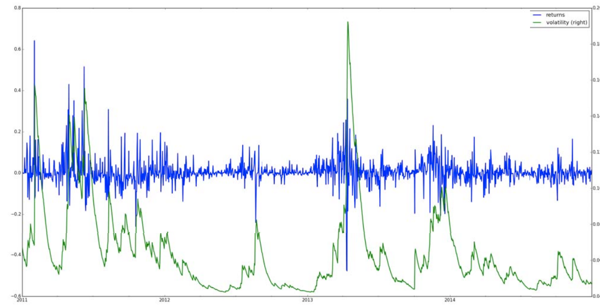

Fig. 3: Daily Bitcoin market price return and volatility. Volatility is estimated from the GARCH(1,1) model. Sample period is from 1/1/2011/to 12/31/2014.

TABLE II: The basic statistics of nodes and edges in the daily Bitcoin flow network

The number of nodes and edges in the daily Bitcoin flow network

	Max	Min	Mean	std
The number of nodes	154673	6273	46628	32396
The number of edges	242916	8118	75601	50975

TABLE II shows the basic statistics for the number of nodes and edges in a daily Bitcoin transaction network.

B. Network Complexity

In this section, we introduce a couple of the network complexity measures which are originally introduced by theoretical ecologists. First, we introduce the concepts of the connectivity

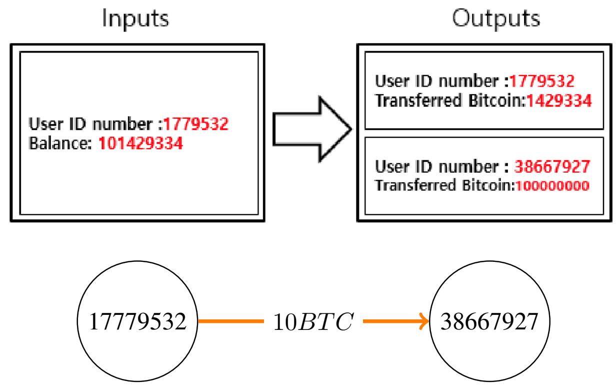

Fig. 4: The upper diagram shows the processed transaction data with the user ID number and amount of transaferred BTC between the user IDs. The lower diagram is the part of Bitcoin transaction network reflecting the transaction in the upper diagram.

and the number of roles of a network. Second, we examine the complexity measures from information theory. Interestingly we find that the connectivity and the number of roles in a network is equivalent to the complexity measures from the information theory.

1) Connectivity and the Number of Roles: In Ulanowicz [26], the degree of the network complexity is quantified by the concepts of connectivity and the number of roles of a network. First, these two concepts are defined for an unweighted network and are then extended into a weighted one. In the unweighted network, the connectivity of a network is defined as the average number of edges connected to a node. Let N, F, and C be the number of nodes, the number of edges, and the connectivity of a network respectively. Then the connectivity can be defined as $C = F/N$. And the number of roles is defined as the ratio of the connectivity of a network to the number of nodes in the network. Let $R = C/N = N^2/F$ be the number of roles in a network. A role is, loosely speaking, a specialized function defined as a group of nodes that take their input from one source and pass them to a single destination. In Fig. 5, the left network has the connectivity of 1 with 3 roles, $C = 1$ and $R = 3$. Intuitively, this makes sense because all three nodes are playing a unique role in the network, taking its inputs from one source and passing them onto another. The right network in Fig. 5 has the connectivity of 3 with 1 role, $C = 3$ and $R = 1$. This also makes sense because the 3 nodes are not distinguishable from each other, because they receive from every other node and give to every other node in the network. Even though we have some sense about the two complexity measures from a special case, they do not make much sense in general cases for an unweighted network. Ulanowicz [25] extends theses complexity measures into a weighted network. A weighted network is a network whose edges among nodes have weights assigned to them. Because the weight on an edge means the relative importance of an edge, the weight on an edge plays an important role in defining the connectivity of a node and consequently the connectivity of a network. Thus we begin with the connectivity of a node.

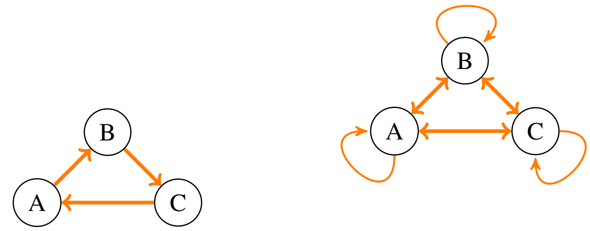

Fig. 5: The left diagram shows the network with the connectivity of 1 and 3 roles. The right diagram shows the network with the connectivity of r and 1 role.

In Fig. 6, the node *i* receives 50BTC from each of the node *h1* and the node *h2* and sends 99 BTC and 1 BTC to the node *j1* and the node *j2* respectively. Intuitively, for the node *i*, the input connectivity should be 2 and the output connectivity should be close to 1 because 99% of total outflows flows out through a single edge. Therefore, we need to weight each edge pointing in(out) by the relative size of the weight on each edge to the sum of weights on the edges pointing in(out). Among various weighting schemes [25], the weighted geometric mean is used and is proven to be the only one logically consistent with the unweighted network to extend complexity measures into the weighted network. Let T_{ij} denote the magnitude of a flow from a node *i* to a node *j*. Let $T_{i\cdot} = \sum_j T_{ij}$ denote the total magnitude of outflows of a node *i*, $T_{\cdot j} = \sum_j T_{ij}$ denote the total magnitude of inflows of a node *j*, and $T_{\cdot\cdot} = \sum_{i,j} T_{ij}$ denote the total magnitude of flows within a network. $T_{\cdot\cdot}$ is known as the total system throughput (TST). And we call $T_{i\cdot}$ and $T_{\cdot i}$ the output and input throughputs of a node *i* respectively.

In- and Out- connectivity for vertex *i* are defined as following:

$$IC(i) = \prod_h \left(\frac{T_{\cdot i}}{T_{hi}}\right)^{T_{hi}/T_{\cdot i}}, \quad OC(i) = \prod_j \left(\frac{T_{i\cdot}}{T_{ij}}\right)^{T_{ij}/T_{i\cdot}} \tag{1}$$

And the connectivity of a vertex is the weighted geometric mean of in- and out- connectivity of the vertex. The weight for in-connectivity is the ratio of total inputs to total vertex throughputs, and similarly for the weight to out-connectivity.

$$\begin{aligned} C(i) &= IC(i)^{\frac{T_{\cdot i}}{T_{\cdot i}+T_{\cdot i}}} OC(i)^{\frac{T_{i\cdot}}{T_{\cdot i}+T_{i\cdot}}} \\ &= \prod_h \left(\frac{T_{\cdot i}}{T_{hi}}\right)^{T_{hi}/T_{i\cdot}+T_{\cdot i}} \prod_j \left(\frac{T_{i\cdot}}{T_{ij}}\right)^{T_{ij}/T_{i\cdot}+T_{\cdot i}} \end{aligned} \tag{2}$$

The effective network connectivity is the weighted geometric means over all vertex connectivity. We weight the contribution to the total connectivity from each vertex by its weight in the system. The weight for connectivity of a vertex is the ratio of total vertex throughput to twice the total system throughput.

$$C = \prod_i C(i)^{(T_{i\cdot}+T_{\cdot i})/2T_{\cdot\cdot}} = \prod_{h,j} \left(\frac{T_{hj}^2}{T_{h\cdot}T_{\cdot j}}\right)^{-T_{hj}/2T_{\cdot\cdot}} \tag{3}$$

The effective number of flows of a weighted network is the weighted geometric mean over all edges. The weight for each

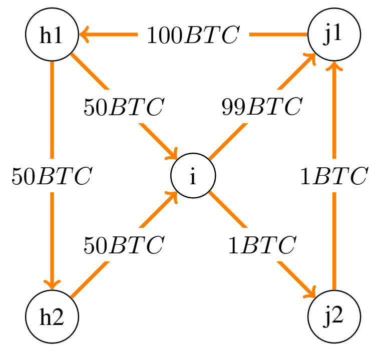

Fig. 6: Weighted Bitcoin transaction network illustrates the need to weight size of each flow and each vertex in network.

TABLE III: Inputs, outputs, and total connectivity. weight for each vertex. Graph Connectivity. Number of nodes, flows, and roles.

node	In. Conn.	Out. Conn.	Conn.	Weight	Estimated measures	
i	2	1.0576	1.4544	0.2849	Eff. # of nodes	3.9254
h1	1	2	1.4142	0.2849	Eff. # of flows	4.8593
h2	1	1	1	0.1425	Eff. Connectivity	1.2379
j1	1.0576	1	1.0284	0.2849	Eff. # of roles	3.171
j2	1	1	1	0.0028		

edge is the ratio of the size of the weight on the edge to TST.

$$F = \prod_{h,j} \left(\frac{T_{..}}{T_{hj}} \right)^{T_{hj}/T_{..}} \tag{4}$$

The effective number of vertices is calculated from the connectivity formula, $C = F/N$.

$$N = \prod_{h,j} \left(\frac{T_{..}^2}{T_{h\cdot}T_{\cdot j}} \right)^{T_{hj}/T_{..}} \tag{5}$$

The effective number of roles is calculated from the formula, $R = N/C$.

$$R = \prod_{h,j} \left(\frac{T_{hj}T_{..}}{T_{h\cdot}T_{\cdot j}} \right)^{T_{hj}/T_{..}} \tag{6}$$

In TABLE III, the above mentioned network metrics for the weighted network in Fig. 6 are calculated. Vertex *i* has 2 of the in-connectivity and its out-connectivity is close to 1. And vertex *j2* only receives and sends 1BTC so that its weight is negligible.

2) Complexity Measure from Information Theory: Information theory (IT) is one approach to quantify information, and its application can be invoked to measure the relative degrees of the constraint and flexibility inherent in a system. IT begins with the Boltzmann's surprisal to quantify "Which is missing". The Boltzmann's surprisal (s) of an event is estimated from $s = -klog(p)$ where s is one's surprisal at seeing an event that occurs with probability p. However, Boltzmann's surprisal measures the absence, not presence of an event. For the small probability that an event occurs, the corresponding surprisal s has large magnitude, which means the event in question is not likely to occur. The product of the measure of presence of an event i (p_i) by the magnitude of its absence (s_i) yields a quantity that represents the indeterminacy (h_i) of the event in question.

$$h_i = -kp_i log(p_i) \tag{7}$$

It is helpful to reinterpret equation [7] as it relates to evolutionary change and sustainability. h_i represents the capacity for event i to be a significant player in a system change or evolution. Regarding the entire ensemble of events, one can aggregate all the indeterminacies to create a metric of the total capacity of the system subject to change.

$$H = \sum_i h_i = -k \sum_i p_i log(p_i) \tag{8}$$

MacArthur applied the above well known Shannon's information measure to the flows in an ecosystem network as an important tool to define the system diversity and stability [14].

$$H = -k \sum_{h,j} \left(\frac{T_{hj}}{T_{..}} \right) log \left(\frac{T_{hj}}{T_{..}} \right) \tag{9}$$

Later Rutledge et al. were able to decompose MacArthur's index into two complementary terms using the notion of conditional probability [23]. Taking $(T_{ij}/T_{..})$ as the estimate of the unconditional probability that a flow occurs from i to j, $(T_{ij}/T_{\cdot j})$ then becomes the estimator of the conditional probability that any quantum of flow continues on to component j, given that it had originated from component i. This allows H to be decomposed into two parts.

$$H = AMI + H_c \tag{10}$$

where

$$\begin{aligned} AMI &= k \sum_{h,j} \left(\frac{T_{hj}}{T_{..}} \right) log \left(\frac{T_{hj}T_{..}}{T_{h\cdot}T_{\cdot j}} \right) \\ and \quad H_c &= k \sum_{h,j} \left(\frac{T_{hj}}{T_{..}} \right) log \left(\frac{T_{hj}^2}{T_{h\cdot}T_{\cdot j}} \right) \end{aligned} \tag{11}$$

The overall complexity of the flow structure, as measured by MacArthur's index, can be decomposed into two parts, AMI and Hc. AMI is called the average mutual information inherent in the flow structure. AMI measures how orderly and coherently the flows are connected, while H_c measures the residual diversity or freedom in the network. It is interesting to see that the complexity measures developed in IT are equivalent to the connectivity and the number of roles defined in the previous section. MacArthure's index is equivalent to the logarithm of the number of flows in the weighted network. AMI and Hc are equivalent to the logarithm of the number of roles and the connectivity respectively. Equation [12] shows this equivalence between the two complexity measures.

$$\begin{aligned} H = k \cdot log(F) = k \cdot log \left(\frac{F}{C} \cdot C \right) \\ = k \cdot log(R) + k \cdot log(C) = AMI + Hc \end{aligned} \tag{12}$$

We adopt these complexity measures and apply them to the Bitcoin transaction network. The flow is defined as the amount of transferred Bitcoin. The above three quantities tell us the complexity, order and freedom in the whole network, from which we can find out how the diversity or freedom is changing through the entire network.

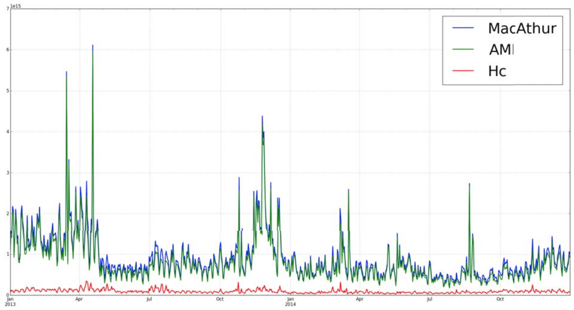

Fig. 7: The three network complexity measures: Scaled MacAthur's index in blue, scaled AMI in green, and scaled Hc in red. All time series range from 1/1/2013 to 12/31/2014.

TABLE IV: The basic statistics of the complexity measures

	Max	Min	Mean	S.D.
S. MacAthur	$3.046 \cdot 10^{16}$	$1.564 \cdot 10^{14}$	$1.379 \cdot 10^{15}$	$2.274 \cdot 10^{15}$
Scaled AMI	$3.031 \cdot 10^{16}$	$1.209 \cdot 10^{14}$	$1.283 \cdot 10^{15}$	$2.260 \cdot 10^{15}$
Scaled Hc	$3.477 \cdot 10^{14}$	$2.376 \cdot 10^{13}$	$9.614 \cdot 10^{13}$	$4.368 \cdot 10^{13}$

In addition to the above network complexity measures, we take account of the amount of transferred Bitcoin during a day. We scale the above equality among the complexity measures by the amount of transferred Bitcoin during a day and have the additional set of three complexity measures. The followings are the equations of the scaled complexity measures.

$$Scaled \quad MacAthur \quad index = \sum_{i,j} T_{ij} log\Big(\frac{T_{ij}T_{..}}{T_{i.}T_{.j}}\Big) \quad (13)$$

$$Scaled \quad AMI = -\sum_{i,j} T_{ij} log\Big(\frac{T_{ij^2}}{T_{i.}T_{.j}}\Big) \quad (14)$$

$$Scaled \quad Hc = -\sum_{i,j} T_{ij} log\Big(\frac{T_{ij}}{T_{..}}\Big) \quad (15)$$

The scaled AMI measures how well the network is progressing and how the medium is flowing within the network. The scaled Hc measures how diversified the source and the destination of the flow is. The equality among the three complexity measures suggests that the increase in the scaled AMI comes at the expense of the scaled Hc. In Fig. 7, we plot the time development of three complexity measures.

V. Market Variables Prediction Model

In this section, we use the complexity measures to forecast Bitcoin market price and volatility. We employ vector autoregression (VAR) to observe the evolution and the interdependencies between the network complexity and market price. In order to test predictability of these network variables, we test Granger-Causality relationship between market variables such as return and volatility and network complexity defined in section IV. We perform Granger causality analysis for each pair of market variable and complexity measure and discuss the results.

TABLE V: The result of the augmented Dickey Fuller(ADF) test for a unit root

Time series	ADF test statistics	95% critical value	Order of Integration
Scaled MacAthur's Index	-3.134	-2.8642	I(0)
Scaled AMI	-3.2297	-2.8642	I(0)
Scaled Hc	-3.4365	-2.8642	I(0)
Return	-9.7597	-2.8642	I(0)
Volatility	-3.7102	-2.8642	I(0)

TABLE VI: Selected lag order for the VAR model of each pair of the market variable and complexity measure.

Complexity measure	Scaled MacAthur		Scaled AMI		Scaled Hc	
Market variable	return	volatility	return	volatility	return	volatility
Lag order	18	9	17	7	15	2

A. Vector Autoregression

Vector autoregression was introduced as a technique to characterize the joint dynamic behavior of a collection of variables without requiring strong restrictions of the kind needed to identify underlying structural parameters. In our analysis, a VAR system contains a pair of a market variable and a complexity measure. Each variable is expressed as a linear function of n lags of itself and the others, plus an error term. Our VAR system takes the following form:

$$\begin{aligned} M_t &= a + \sum_{i=1}^{n} \alpha_{1i} M_{t-i} + \sum_{i=1}^{n} \beta_{1i} C_{t-i} + \epsilon_{1t} \\ C_t &= b + \sum_{i=1}^{n} \alpha_{2i} M_{t-i} + \sum_{i=1}^{n} \beta_{2i} C_{t-i} + \epsilon_{2t} \end{aligned} \quad (16)$$

where M_t and C_t are market variable and complexity measure respectively, and ϵ_{1t} and ϵ_{2t} are the noise terms.

Before implementing VAR, all time series are tested for stationarity with Augmented Dickey Fuller (ADF) test. TABLE V shows that MacArthur's index and AMI are stationary at the first order difference and the other time series are stationary at the ordinary level I(0).

B. Order Selection and Model Estimation

To check for the adequacy of the estimated vector autoregression models, vector autoregression lag order selection criteria are used to choose the appropriate model orders. We employ Akaike Information Criteria (AIC). TABLE VI shows the lag order that has the minimum AIC value for each combination of market variable and complexity measure. After choosing lag order to define VAR model, we estimate all combinations of market variable and complexity. In TABLE VII, we list the estimated coefficients of VAR model for a pair of volatility and system overhead. Most of coefficients are statistically significant.

VI. Discussion

A. Granger-Causality Test Results

The Granger causality test is a hypothesis test to determine whether one time series is useful in forecasting another. The complexity of the Bitcoin network C_t is said to Granger cause

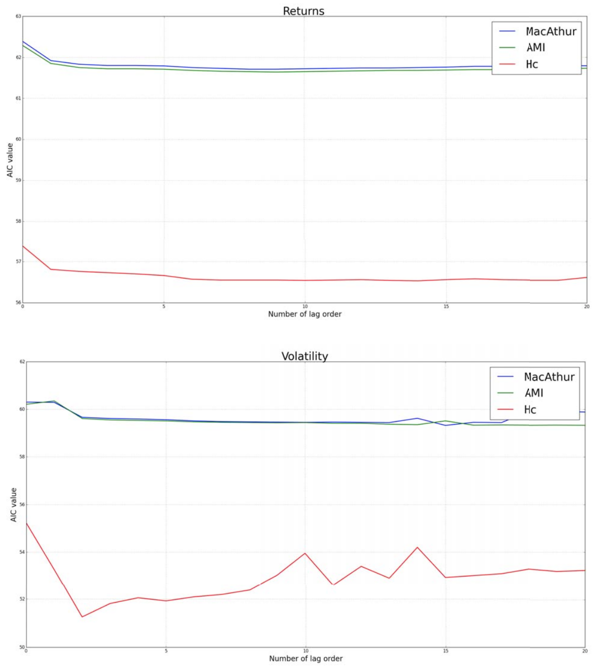

Fig. 8: AIC value over time lag order. In upper panel, AIC values from the VAR model for the pair of Bitcoin return and the network complexity measure. In lower panel, AIC values from the VAR model for the pair of Bitcoin volatility and the network complexity measure.

TABLE VII: The result of VAR estimation for the pair of volatility and the scaled Hc via the maximum likelihood. L1 of a time series means time lag 1, and L2 means time lag 2. * and ** denote the significant parameters at the 95 and 99 percent significance level, respectively.

	Estimated Coefficient for the equation of the scaled Hc	Estimated Coefficient for the equation of volatility
const	26063394205790**	-0.001531**
L1.the scaled Hc	0.672516**	-0.000000**
L1.volatility	96542732256270	0.480548**
L2.the scaled Hc	-0.060770	0.000000**
L2.volatility	99015491314371	0.492765**

the market variable M_t if it can be shown that the values of C_t provide a statistically significant information about the future values of M_t. We conduct the bivariate Granger-Causality test for all pairs between the network complexity measures and market varialbes. In table VIII, the test hypothesis and the result of the bivariate Granger-Causality tests are described. Overall, we find that there exits a strong Granger causality relationship between the scaled Hc and the market variables. Judged by the F-stats, the 95% critical value, and the p-value, we conclude that the Granger causal relationship is mutual between the scaled Hc and return as well as between the scaled Hc and volatility. The interpretation of this result is that the residual diversity or flow freedom measured by the Hc complexity measure Granger causes Bitcoin market return and volatility change, and this causal relationship also has reversal effect. We will further interpret the mutual impact through the analysis of the VAR output next.

TABLE VIII: The result of th bivariate Granger Causality Test for each pair of the market variables and complexity measures. Note: 'Mac. idx.' stands for 'MacAthur's index', and Vol. stands for 'volatility' of Bitcoin price, and Rej. stands for 'reject null hypothesis'.

Null Hypothesis	F-stats	95% critical value	p-value
Scaled Mac. id. doesn't Granger cause return. Return doesn't Granger cause scaled Mac. idx.	0.5714 0.0000	1.6088	0.922 1.000
Scaled Mac. id. doesn't Granger cause vol.(Rej.) Vol. doesn't Granger cause scaled Mac. idx.	3.0234 0.0000	1.8842	0.001** 1.000
Scaled AMI doesn't Granger cause return. Return doesn't Granger cause the scaled AMI.	0.5920 0.0000	1.6276	0.900 1.000
Scaled AMI doesn't Granger cause vol. Vol. doesn't Granger cause the scaled AMI.	0.5799 0.0001	2.0138	0.773 1.000
Scaled Hc doesn't Granger cause return.(Rej.) Return doesn't Granger cause scaled Hc.(Rej.)	2.3730 3.8257	1.6711	0.002** 0.000**
Scaled Hc doesn't Granger cause vol.(Rej.) Vol. doesn't Granger cause scaled Hc.(Rej.)	96.7228 6.6701	2.9998	0.000** 0.001**

B. Impulse Response Function Analysis

Now we analyze the impulse response function (IRF) of VAR analysis output. In general, an impulse response refers to the reaction of any dynamic system in response to some external change. In Fig. 9, we plot IRF from the VAR system containing the Bitcoin market variables (return and volatility) and the scaled Hc. We find that when there is an impulse on the scaled Hc, volatility increases sharply in one day and then keeps increasing afterward, and overall the scaled Hc keeps positive response. Similarly, when there is an impulse on the scaled Hc, the return increases sharply in one day and then continues to increase with a rather flat rate. However, on the third day, there exists a return to Hc reverse impact. It seems evident that the Grander causal relationship between the return and the scaled Hc is not monotonic as it is the case between the volatility and the scaled Hc.

VII. Conclusion

We empirically study the Bitcoin transaction data from its inception to 12/31/2014, and we propose two network flow measures to quantify the dynamics of the entire Bitcoin transaction system. We constructed a VAR model to examine the dynamic relationship between network flow complexity and Bitcoin market variables. And we find that the network flow complexity Hc is statistically significant in forecasting the Bitcoin market return and volatility. After we take the total amount of transferred Bitcoin during a day into our consideration, we find that the causality relation between the volatility and the Hc complexity is more pronounced. Furthermore, the causality between this scaled complexity measure and the return is also proved to be statistically significant. We argue that the residual diversity or the freedom of Bitcoin network flow scaled by the total system throughput can significantly improve the predictability of Bitcoin market variables.

References

[1] D. Baird, R.E. Ulanowicz, R.E. *The seasonal dynamics of the Chesapeake Bay ecosystem.* Ecol Monogr 1989, 59, 329?364.

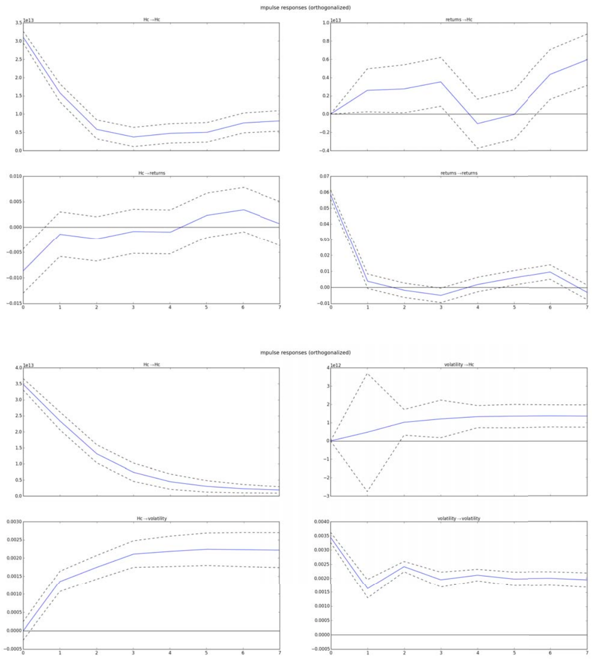

Fig. 9: The upper 4 graphs shows the impulse response between Bitcoin price return and the scaled Hc. The graph in the upper left conner indicates the impulse response of the scaled Hc to the scaled Hc, one in the lower left conner shows the impulse response of Bitcoin return to the scaled Hc, one in the upper right conner shows the impulse response of the scaled Hc to Bitcoin return, and one in the lower right conner shows the impulse response of Bitcoin return to volatility. The lower 4 graphs show the impulse response between Bitcoin price volatility and the scaled Hc.

[2] C. Belady, K.M. Kelkar, and S.V. Patankar, *Improving productivity in electronic packaging with flow network modeling* (FNM). Electronics Cooling, 1999.

[3] M. Benzi and C. Klymko, *A matrix analysis of different centrality measures*. arXiv. Retrieved July 11, 2014.

[4] P. Bonacich, *Power and Centrality: A Family of Measures*. American Journal of Sociology (University of Chicago Press) 92: 1170?1182, 1987

[5] S. P. Borgatti, *Centrality and Network Flow*. Social Networks (Elsevier) 27: 55?71, 2005.

[6] S. P. Borgatti, and M. G. Everett, *A Graph-Theoretic Perspective on Centrality*. Social Networks (Elsevier) 28: 466?484, 2006.

[7] L.R. Ford, D.R. Fulkerson, *Maximal flow through a network*. Can J Math 1956, 8, 399?404.

[8] L. Freeman, *A Set of Measures of Centrality Based on Betweenness,* Sociometry, vol. 40, no. 1, pp. 3541, 1977.

[9] D. Haydon, *Pivotal assumptions determining the relationship between stability and complexity: An analytical synthesis of the stability-complexity debate*. Am Nat 1994, 144, 14?29.

[10] S.A. Kauffman, *Antichaos and adaptation*. Sci Am 1991, 265, 78?84.

[11] S.A. Kauffman, *The Origins of Order*; Oxford University Press: New York,1993.

[12] D. Kondor, M. Posfai, I. Csabai, and G. Vattay, *Do the rich get richer? An empirical analysis of the Bitcoin transaction network* PLoS ONE. 9(2), 2014

[13] W. Leontief, *The Structure of the American Economy*, 1919?1939. 2nd ed. Oxford University Press, New York, 1951.

[14] R. MacArthur, *Fluctuations of naimal populations and a measure of community stability*, Ecology, 553-536,1995

[15] R.M. May, *Will a large complex system be stable?*, Nature, 238, 413?414, 1972.

[16] M.E.J. Newman, *The mathematics of networks. In: The New Palgrave Dictionary of Economics*, 2nd edition (ed. Durlauf SN, Blume LE). Palgrave Macmillan, Basingstoke, 2008.

[17] M.E.J. Newman, *Networks: An Introduction*. Oxford, UK: Oxford University Press, 2010.

[18] M. Ober, S. Katzenbeisser, and K. Hamacher, *Structure and Anonymity of the Bitcoin Transaction Graph*, Future Internet 5(2), pp. 237-250, 2013

[19] L. Peusner, *Studies in Network Thermodynamics*; Elsevier, Amsterdam, 1986.

[20] F. Reid, and M. Harrigan, *An Analysis of Anonymity in the Bitcoin System*, Security and Privacy in Social Networks, Springer, pp. 197?223, 2012

[21] M. Rejmanek, and P. Stary, *Connectance in real biotic communities and critical values for stability of model ecosystems*. Nature 1979, 280, 311?313.

[22] D. Ron, and A. Shamir, *Quantitative Analysis of the Full Bitcoin Transaction Graph*, In Proceeding of the 17th International Conference on Financial Cryptography and Data Security in Okinawa, Japan, 2013.

[23] R.W. Rutledge, B.L. Basorre, and R.J. Mulholland, *Ecological stability: an information theory viewpoint*. J Theor Biol, 57, 355?371, 1976.

[24] R.E. Ulanowicz, *Growth and Development: Ecosystem Phenomenology*; Springer-Verlag: New York, 1986.

[25] R.E. Ulanowicz, *Quantitative methods for ecological network analysis*, Computational Biology and Chemistry 28.5321-339 2004

[26] R. E. Ulanowicz and A. C. Zorach, *Quantifying the complexity of flow networks: how many roles are there?* Complexity,8(3) 68-76, 2003

[27] D.J. Watts and S.H. Strogatz, *Collective Dynamics of ?Small-World? Networks*, Nature 393, 440-442, 1998.

[28] F. Wulff, J.G. Field, K.H. Mann, *Network Analysis in Marine Ecology. Coastal and Estuarine Studies Series*; Springer-Verlag, Berlin, 1989.