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 returns and volatility. We find that complexity of Bitcoin transaction network is significantly correlated with Bitcoin market volatility. More specifically we document that the popularity of Bitcoin gauged from total system throughput can significantly improve the predictability of Bitcoin market returns and volatility using network flow complexity measures.**

Keywords: Bitcoin; Network analysis; Network flow; Entropy; Network complexity

I. Introduction

Bitcoin is a peer-to-peer decentralized crypocurrency system proposed by Satoshi Nakamoto[20]. Unlike a traditional currency, Bitcoin relies on 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 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.

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 vertices of the network, and the edges represent transactions among these participants. Bitcoin transaction network therefore represents all the activities within the Bitcoin market, and they are recorded in the 'block chain'. 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. 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. 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 size, with sometimes extraordinary differences [26]. And hence the question becomes "What characterizes the flow among vertices?"

In this study, we investigated two weighted network complexity measures to quantify the movements of network flows as a time-series, and we aim to reveal the inherent interrelationships between these network complexity measures and the Bitcoin market variables. The central measures used are the numbers of flows and vertices, connectivity in flows per vertex, and the number of roles as measured by vertices 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 III presents Bitcoin market and transaction data. It explains the mechanism of the 'block chain' transactions. Section IV describes the methodology for the construction of Bitcoin transaction flow networks and the weighted network complexity measures. Section V introduces market variable prediction model using the network complexity measures. Section VI discusses the performance of the predictive model, and section VII concludes the findings and point to future research directions.

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 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 centralities count walks which start/end from a given vertex. 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 vertex. The typical example is Freedman's betweenness centrality, the number of shortest paths which pass through a given vertex [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 first 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 the anonymity in Bitcoin network and analyzed the topology of two types of networks: the transaction network and user network [20]. Networks were constructed from the time interval ranging between Bitcoin's inception1/3/2009 and 5/13/2012. To construct a user network, they proposed a data pre-processing to identify vertices in the user network. To improve identification accuracy in the user network, the authors used external information in which Bitcoin users reveal their public addresses online. They showed that it is possible to map Bitcoin public address to IP address and link to 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 majority of minted Bitcoin are not in use and that a large number of tiny transactions exist, and there are also hundreds of transactions that moved more than 50,000 Bitcoin(BTC). And they analyzed these big transaction networks and found that most of these transactions are successors of the initial ones. Another interesting finding is that transaction flows reveal some 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 the study. Overall we collected a comprehensive historical transaction and market price dataset. Both of which 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. All historical Bitcoin transactions are recorded in 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 Bitcoin's inception until 12/31/2014. Here, a Bitcoin transaction is defined as the change of ownership of Bitcoin between the Bitcoin's public addresses. The transaction data is comprised of two parts, inputs and outputs. An input of Bitcoin transaction specifies Bitcoin public address and the amount of Bitcoin stored in the public address. An output specifies Bitcoin public address and the amount of Bitcoin transferred into the public address. Bitcoin transaction can have multiple inputs and multiple outputs. The Fig. 1 shows the real transaction data with 2 inputs and 2 outputs. The amount of Bitcoin is represented in Satoshi. One Bitcoin is equivalent to 10^8 Satoshi.

B. Bitcoin Market

Fig. 2 exhibits the time series of 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 US$1000 all-time high on November 28, 2013 and fell to around $400 in April 2014. The price quickly

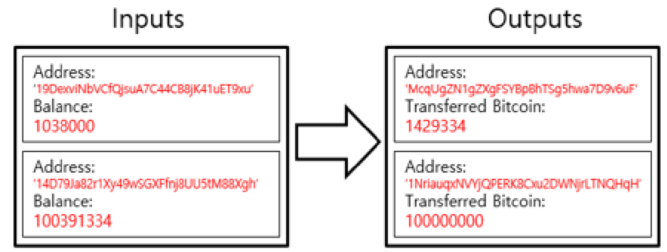

Fig. 1: Bitcoin transaction # 2642247 generated at 1/1/2012 6:22:01

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

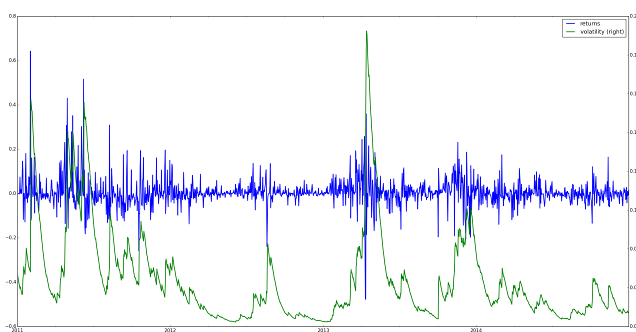

Fig. 3: Daily Bitcoin market returns and volatility estimated from GARCH(1,1). Sample period is from 1/1/2011/to 12/31/2014.

rebounded, returning to \$600 and steadily declined to not much more than \$200.

Daily Bitcoin return is defined as the difference between log of opening and closing index prices. We use GARCH model to estimate the volatility of Bitcoin market price. In Fig. 3, we plot daily Bitcoin return and volatility estimated from GARCH(1,1). The basic statistics for daily return and volatility are summarized in TABLE I.

IV. Methodology

A. Bitcoin Network

We construct a Bitcoin transaction network for each day during the sample period. Network is composed of edges and vertices. We follow the network construction method used by Reid and Harrigan [20]. In Reid and Harrigan [20], a vertex of network is defined as a group of Bitcoin public addresses which are assumed to be owned by the same entity. Every Bitcoin address in a vertex should be involved in the input part of a Bitcoin transaction with the other addresses in the same vertex. 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 Bitcoin transaction. To identify the ownership of public addresses, we implement the Union-Find algorithm to 59,151,261 public addresses, which results in over 5 million entities. And we then assign a user identification number to each entity. An edge of network indicates that there is a transaction between two entities. Edge has directionality from the entity on the input side to the entity on the output side. Since there could be multiple entities on the output side of a transaction, multiple edges could occur from a single transaction. If the entity in the output is the same as the entity in the input, we then will not consider this a self-loop. A weight on an edge indicates the amount of transferred Bitcoin from a transaction. Since there could be multiple transactions from one vertex to another during a day, the weight on edge is the sum of weights from the multiple transactions. Fig. 4 shows the processed Bitcoin transaction data with user identification number, and the part of transaction network is reflected from the processed transaction data. Before transaction #2642247, user #1777953210 has 10.1429334 BTC. 10 out of 10.1429334 BTC, and it is then transferred from user #17779532 to user #38667927 and the change of 0.1429334 BTC is kept by user #17779532.

TABLE I: Basic statistics of Market variables. Returns is a log difference of Bitcoin prices, and volatility is estimated from GARCH(1,1) model. Sample period is from 1/1/2011 to 12/31/2014.

	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

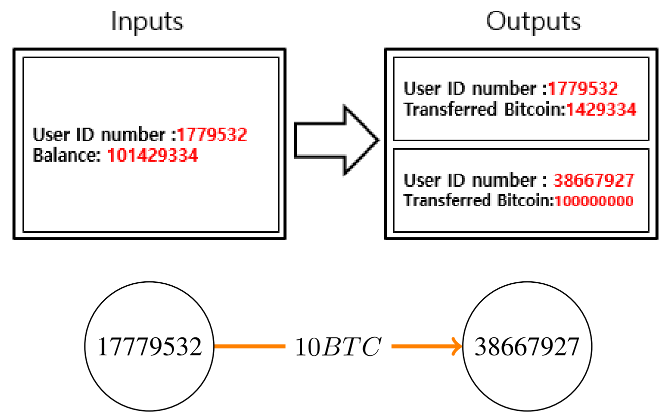

Fig. 4: User identification number is assigned to a public address and is reflected in network from the processed transaction data.

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

TABLE II: Basic Statistics for Network Components

The number of vertices and edges in network

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

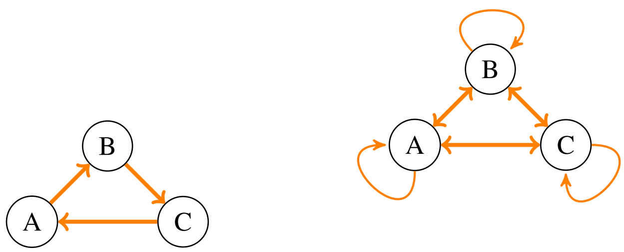

Fig. 5: An unweighted network with 3 roles and 1 role

B. Network Complexity

To measure the weighted network flows in Bitcoin networks, we adopt the connectivity and the number of roles in network introduced by Ulanowicz [26]. In addition, we also examine the complexity measures from information theory. Interestingly we find that the connectivity and the number of roles in network is equivalent to the complexity measures from information theory. In this section, we describe the defintition and mathematical derivation of these complexity measures.

1) Connectivity and the Number of Roles: In Ulanowicz [26], complexity measures called the connectivity and the number of roles are introduced. First, these two measures are defined for unweighted networks.An unweighted network is a collection of vertices and edges with equal weight. In an unweighted network, the connectivity is defined as the average number of edges per vertex. Let N, F, and C be the number of vertices, the number of edges, and the connectivity of a network respectively. Then the connectivity is $C = F/N$. And the number of roles is defined as the ratio of connectivity to the number of vertices in network. Let $R = C/N = N^2/F$ be the number of roles. A role is, loosely speaking, a specialized function: it is a group of vertices that takes its inputs from one source and passes them to a single destination. In Fig. 5, the left network has a connectivity of 1 and 3 roles, $C = 1$ and $R = 3$. Intuitively, this makes sense because all three vertices 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 a connectivity of 3 and 1 role, $C = 3$ and $R = 1$. This also makes sense because the 3 vertices are not distinguishable from each other, because they receive from every other vertex, and give to every other vertex in the network. Even though we have some sense about the two complexity measures from a special case, they do not make sense in general cases in an unweighted network. Ulanowicz [25] extends theses complexity measures into the weighted network. A weighted network has different sizes of weights on edges, and we must take the size of weight into consideration. The size of weight on edge plays an important role in defining the connectivity of a vertex and, in turn, the connectivity of a network. We begin with the connectivity of a vertex. In Fig. 6, vertex *i* receives 100BTC equally from vertex *h1* and vertex *h2* and sends 99 BTC and 1 BTC to vertex *j1* and vertex *j2* respectively. Intuitively, the input connectivity of vertex *i* 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, in Ulanowicz [25], the weighted geometric mean is used and proved to be the only one logically consistent way with the unweighted network to extend complexity measures into the weighted network. Before observing the network metrics formula, let T_{ig} represent the flow from vertex i to vertex j. Let $T_{i\cdot} = \sum_j T_{ij}$ be the total flow out of vertex *i*, and $T_{\cdot j} = \sum_j T_{ij}$ be the total flow into vertex *j*. Let $T_{\cdot\cdot}$ is known as the total system throughput (TST), and we call $T_{i\cdot}$ and $T_{\cdot i}$, respectively, the output and input throughputs of a vertex *i*.

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

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

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

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

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

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

In TABLE III, the above mentioned network metrics for the weighted network in Fig. 6 are calculated. Vertex *i* has 2 of

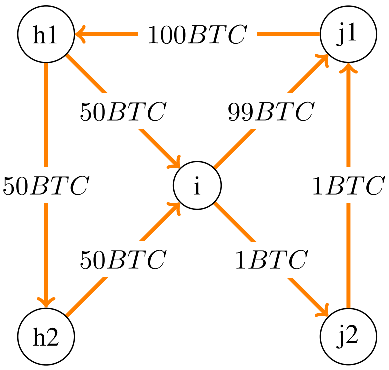

Fig. 6: Weighted Bitcoin transaction network illustrates the need to weight size of each flow and each vertex in network.

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

Vertex	In. Conn.	Out. Conn.	Conn.	Weight	Estimated measures	
i	2	1.0576	1.4544	0.2849	Eff. # of vertices	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		

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 constraint and flexibility inherent in a system. IT begins with Boltzmann's surprisal to quantify "Which is missing". Boltzmann's measure called surprisal is $s = -klog(p)$ where s is one's surprisal at seeing an event that occurs with probability p. However, Boltzmann's surprisal measures the absence, not presence. 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 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].

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

TABLE IV: Basic statistics of complexity measures

	Max	Min	Mean	S.D.
MacArthur's Index	3.7017	-3.6523	0.004673	0.6915
AMI	12.261126	5.920743	9.678522	1.018448
Hc	3.420247	0.045518	1.212666	0.611452
Dev. capacity	$3.046{\cdot}10^{16}$	$1.564{\cdot}10^{14}$	$1.379{\cdot}10^{15}$	$2.274{\cdot}10^{15}$
Overhead	$3.477{\cdot}10^{14}$	$2.376{\cdot}10^{13}$	$9.614{\cdot}10^{13}$	$4.368{\cdot}10^{13}$
Ascendency	$3.031{\cdot}10^{16}$	$1.209{\cdot}10^{14}$	$1.283{\cdot}10^{15}$	$2.260{\cdot}10^{15}$
A/D.C.	0.99533	0.6805	0.889904	0.054705

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, $(T_{ij}/T_{.j})$ 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

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

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 gauges how orderly and coherently the flows are connected. And H_c gauges the residual diversity or freedom in the network. It is interesting that complexity measures developed in IT are equivalent to 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 connectivity. The Eqn. [12] shows this equivalence between the two complexity measures.

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

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.

In addition to the above network complexity measures, we take the total amount of transferred Bitcoin within a network into consideration. We define three complexity measures considering the total flows within the networks: ascendency, system overhead, and system development capacity. Ascendency of network(A) was formulated to combine the total activity of AMI, which means that its organization in the sense of how effectively component processes are linked. Ascendency measures how well the network is progressing and how the medium is flowing within the network. Scaling MacArthur's

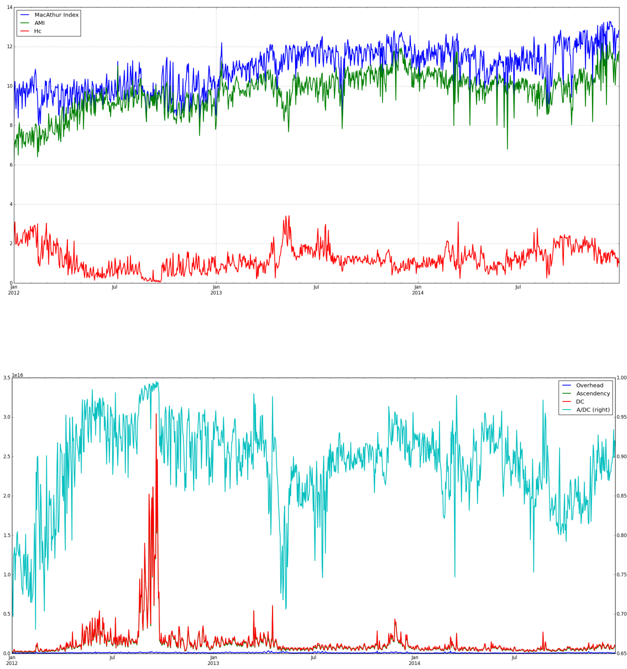

Fig. 7: Network complexity measures. In the upper panel, MacArthur's index, AMI, and Hc are plotted. In the lower panel, system overhead, development capacity, ascendency, and the ratio of ascendency to capacity are plotted. All time series range from 1/1/2012 to 12/31/2014.

index and Hc, we have system development capacity and overhead (Φ). The equivalence relation among three measures is $C = A + \Phi$. This relation suggests that the increase in ascendency comes at the expense of overhead. The portion of ascendency in capacity becomes a convenient measure of the degree of system order.

$$A = \sum_{i,j} T_{ij} log\Big(\frac{T_{ij}T_{..}}{T_{i.}T_{.j}}\Big) \tag{13}$$

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

$$C = -\sum_{i,j} T_{ij} log\Big(\frac{T_{ij}}{T_{..}}\Big) \tag{15}$$

In Fig. 7, we plot the time development of our 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) method to model 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: Augmented Dickey Fuller(ADF) test for a unit root

Time series	ADF test statistics	95% critical value	Order of Integration
MacArthur's Index	-14.2465	-2.8642	I(1)
AMI	-15.0813	-2.8642	I(1)
Hc	-4.0836	-2.8642	I(0)
Development Capacity	-3.134	-2.8642	I(0)
Overhead	-3.4365	-2.8642	I(0)
Ascendency	-3.2297	-2.8642	I(0)
Asc./D.C.	-4.0202	-2.8642	I(0)
return	-9.7597	-2.8642	I(0)
volatility	-3.7102	-2.8642	I(0)

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:

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

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 for determining whether one time series is useful in forecasting another. The dynamics of complexity measure C_t is said to Grangerly-cause market variable M_t if it can be shown that

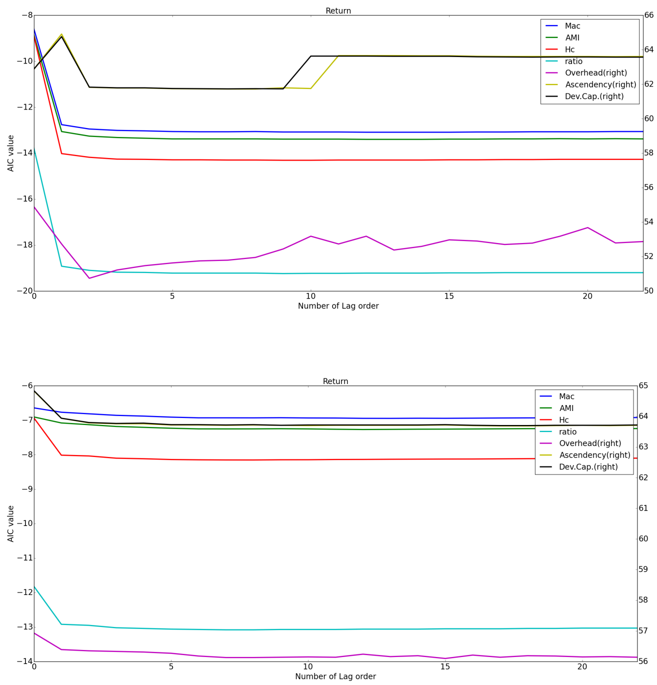

Fig. 8: AIC value for different lag order. In upper panel, AIC values from VAR for Bitcoin return and network complexity measure. In lower panel, AIC values from VAR for Bitcoin volatility and network complexity measure.

TABLE VI: Selected lag order of VAR model for each pair of market variable and complexity measure. r and v stand for return and volatility of Bitcoin respectively. D.C. stands for development capacity. Ascendency(Asc.) Overhead(Oh.) the ratio of Ascendency to development capacity (A/D.C.)

Complexity	Mac		AMI		Hc		D.C.		Asc.		Oh.		A/D.C.	
Market var.	r	v	r	v	r	v	r	v	r	v	r	v	r	v
Selected Lag order	13	14	12	12	8	9	18	9	17	7	15	2	8	9

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

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

those C_t values provide statistically significant information about future values of M_t. In Fig. 9, we draw a diagram to describe the Granger-Causality relationship between network complexity measure and Bitcoin market variables. The strong Granger causality exists between system overhead and market variable. We consider that Bitcoin popularity is assessed by TST. Multiplying AMI and Hc by TST provides a prominent causality relationship between market variable and complexity.

TABLE VIII: Bivariate Granger Causality Test between market variables and complexity measure. D.C. below means system development capacity. A/C means the ratio of ascendency to development capacity

Null Hypothesis	F-stats	95% critical val.	p-value
MacArthur doesn't Granger cause return. Return doesn't Granger cause MacArthur.	0.5469 1.2482	1.7248	0.896 0.238
MacArthur doesn't Granger cause volatility. Vol. doesn't Granger cause MacArthur.	0.8005 1.1563	1.6964	0.669 0.303
AMI doesn't Granger cause return. Return doesn't Granger cause AMI. (REJECT)	0.3817 1.7576	1.75674	0.970 0.050*
AMI doesn't Granger cause vol. Vol. doesn't AMI cause volatility.	0.6289 1.6461	1.7567	0.819 0.073
Hc doesn't Granger cause return. (REJECT) Return doesn't Granger cause Hc.	2.1244 1.0545	1.9427	0.031* 0.392
Hc doesn't Granger cause vol.(REJECT) Vol. doesn't Granger cause Hc.(REJECT)	2.1237 2.3371	1.8842	0.025* 0.013*
D.C. doesn't Granger cause return. Return doesn't Granger cause D.C.	0.5714 0.0000	1.6088	0.922 1.000
D.C. doesn't Granger cause vol.(REJECT) Vol. doesn't Granger cause D.C.	3.0234 0.0000	1.8842	0.001** 1.000
Oh. doesn't Granger cause return.(REJECT) Return doesn't Granger cause Oh.(REJECT)	2.3730 3.8257	1.6711	0.002** 0.000**
Oh. doesn't Granger cause vol.(REJECT) Vol. doesn't Granger cause Oh..(REJECT)	96.7228 6.6701	2.9998	0.000** 0.001**
Asc. doesn't Granger cause return. Return doesn't Granger cause Asc.	0.5920 0.0000	1.6276	0.900 1.000
Asc. doesn't Granger cause vol. Volatility doesn't Granger cause Asc.	0.5799 0.0001	2.0138	0.773 1.000
A/C doesn't Granger cause return.(REJECT) Return doesn't Granger cause A/C.	2.0781 0.7205	1.9427	0.035* 0.674
A/C doesn'tGranger cause vol.(REJECT) Vol. doesn't Granger cause A/C .(REJECT)	1.9435 2.3395	1.8842	0.042* 0.013*

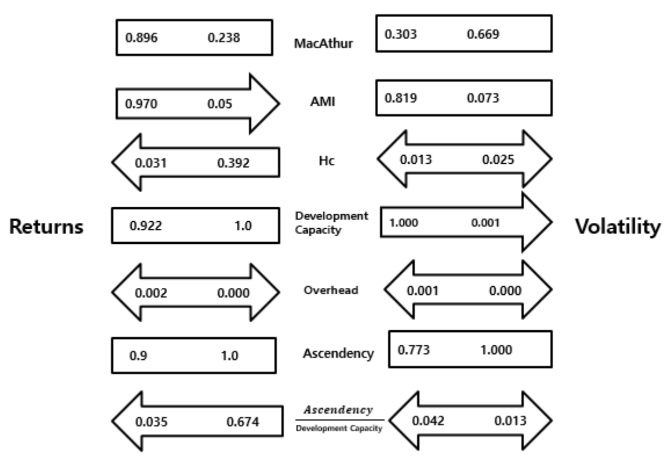

Fig. 9: Diagram for bivariate Granger-Causality test between market and network variables. Directionality of arrow indicates the directionality of causality. For instance, volatility Granger causes the number of roles but the number of roles does not Granger cause volatility. The arrow between Roles and Volatility only has the direction from Volatility to Roles. The numbers in a arrow indicate the p-value of Granger-Causality test.

System disorder(Hc) is helpful to forecast volatility alone. If we consider Bitcoin popularity and Hc is scaled by TST, we have strong causality between market return and complexity measure.

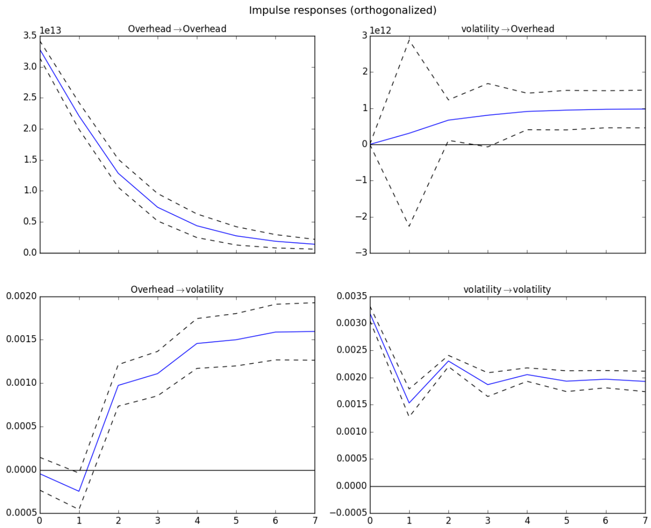

Fig. 10: Impulse Response function between volatility and system overhead. The graph in the upper left conner indicates the impulse response of Overhead to Overhead, one in the lower left conner shows the impulse response of volatility to Overhead, one in the upper right conner shows the impulse response of Overhead to volatility, and one in the lower right conner shows the impulse response of volatility to volatility.

B. Impulse Response Function Analysis

Impulse response function (IRF) of VAR is its output when presented with a brief input signal called an impulse. More generally, an impulse response refers to the reaction of any dynamic system in response to some external change. In Fig. 10, we plot IRF for the pair of volatility and 'Hc' and system overhead. We find that when there is an impulse on system overhead, volatility decreases in one day and keeps increasing afterward, and system overhead keeps positive response after there is an impulse on volatility.

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 network flow complexity is statistically significant in forecasting Bitcoin market volatility. After we take the total amount of transferred Bitcoin during a day into our consideration, we find that the causality relation between volatility and complexity is more pronounced. Furthermore, the causality between the network complexity and return is also proved to be statistically significant. We argue that the popularity of Bitcoin gauged from total system throughput can significantly improve the predictability of Bitcoin market variables using network flow complexity.

References

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

[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.