International Journal of Industrial Engineering & Production Research March 2022 Vol. 33, No. 1: 1-14
DOI: 10.22068/ijiepr.33.1.10

RESEARCH PAPER

# The SCAN method to monitor cryptocurrency transactions

**Fatemeh Elhambakhsh[1] & Kamyar Sabri-Laghaie[2*]**

Received 22 December 2021; Revised 11 January 2022; Accepted 6 February 2022;


**ABSTRACT**
*The fourth industrial revolution has changed our lives by enabling everyone to be interconnected virtually. A trustworthy system is required to secure large volumes of stored data in IoT-based devices. Blockchain technology has led to the safe transfer and storage of data. With this in mind, blockchain-based cryptocurrencies have gained quite a bit of popularity because of their potential for financial transactions. In this regard, monitoring transactions networks is very fruitful to find users' abnormal behaviors. In this research, a novel procedure is used to monitor cryptocurrency transactions networks. To do so, a random binary graph model is used to simulate the transactions between users, and a SCAN method is used to detect the abnormal behaviors in the simulated model. Also, a multivariate exponentially weighted moving average (MEWMA) control chart is used to monitor centrality measures. The probability of signal is used to assess the performance of the SCAN method and that of the MEWMA control chart in distinguishing abnormalities. Then, the procedure is adopted to a Bitcoin transactions dataset.*

**KEYWORDS:** *Blockchain; Social networks monitoring; Cryptocurrency; Bitcoin; MEWMA; SCAN method.*

## 1. Introduction

In recent decades, industry 4.0 has changed our lives by enabling everyone to have interactions with the help of Internet-based technologies [1]. One of the significant concerns among people is data security due to the improvement of the Internet [1]. Therefore, it is of great importance to secure data of numerous smart IoT-based devices [1-3]. A wide range of industry 4.0-based applications, such as manufacturing, finance, healthcare, etc., are concerned with large volumes of data [4, 5]. Thus, a trustworthy system is required to secure such data [1]. The technology of blockchain has brought about new safe ways of data transferring and storing to achieve efficient work process in different industries [6]. It's one of the most important and essential parts of industry 4.0 [5, 7, 8]. For instance, in business sector, blockchain provides improved smart contracts to make easy, safe, as well as secure transactions among users [1].

Nowadays, with the rapid growth of industry 4.0, blockchain-based cryptocurrencies have experienced growing popularity and interest [9], because the advent of blockchain technology brought about easy access to stored financial transaction data [10, 11]. A blockchain is an unchangeable data storehouse in which transactions are prolonged by nodes through the network [1]. Therefore, the most widespread use of cryptocurrency-based blockchain is in the finance sector [12]. The Cryptocurrencies are introduced to our life with Bitcoin [7]. It is the first and most important cryptocurrency that caused a global change which offers safe and better payment options as well as it has the potential to increase financial participation [7]. Blockchain for Bitcoin transactions is a decentralized data transformation technology [13] without involvement of any third party that provides data integrity and security [14]. In this regard, with the high-speed growth of network data, monitoring the attributes of blockchain-based cryptocurrency networks and analyzing their trends can be fruitful in order to achieve numerous information among involving entities.

* Corresponding author: ***Kamyar Sabri-Laghaie***
***sabri@uut.ac.ir***

1. *Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran.*
2. *Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran.*

Networks are complex systems to describe real world problems in different areas of biological, social, cyber, financial, etc. [15, 16]. Especially, a considerable amount of research on cryptocurrency transactions could be done from a network perspective in the financial sector [9]. A network is a complex system by which the relationships among nodes are represented. As mentioned, the transactions of cryptocurrencies on the platform of Blockchain can be modeled as a network. Such transactions among the entities in the network can be monitored to identify unusual behavior of users. This is of great importance, because of lack of regulations in the Bitcoin market, it can cause some kinds of unlawful behaviors [17].

Since the invention of Blockchain and the popularity of cryptocurrencies many researches have focused on analyzing the abnormal behavior of users. In this regard, Wu et al. [9] presented an exhaustive review of state-of-the-art research based on transactions analysis of cryptocurrencies from a network perspective. They summarized techniques for network modeling, network profiling, and transactions network detection. At the end, they provided some challenges and future research directions. To be more specific, they addressed the various models of cryptocurrency networks such as Reid and Harrigan[18], Chen et al.[19], Zhao et al.[20], and Lin et al.[21], to name but a few. In addition, they discussed the analysis of the different properties of cryptocurrency networks, some of which are Chen et al.[22], Ferratti and D`Angelo[23], and Alqassem et al.[23, 24]. Additionally, they reviewed abnormal behavior detection research and methods in cryptocurrency networks, for instance, the studies of Victor[25], Wu et al.[26], and Lin et al.[27]. Liu et al. [28] analyzed data mining methods on cryptocurrency transactions. They provided a complete review on traceability and linkability analysis, collective transaction pattern analysis, and individual user behavior analysis. Anoaica and Levard [29] provided a quantitative analysis on Ethereum network. First, they computed correlation between internal variables and money exchange rate in time. Then, they represented Ethereum graph using user-to-user transactions to find major actors by using different centrality measures of In-Degree/Out-Degree, Betweenness, and Left Eigenvector. Vidal-Tomás [30] analyzed the effect of COVID-19 pandemic on the stability of cryptocurrency networks including, Bitcoin, Ethereum, Ripple, Litecoin. For the network model, they considered individual cryptocurrencies as nodes and the Pearson's correlation coefficient as edges between them. Degree Centrality and Betweenness Centrality are monitored during the pandemic, as well. Bianconi and Agraval [31] proposed a methodology to anticipate the number of transactions for Bitcoin network in the future by computing total number of edges and extracting different features using machine learning methods. Aspembitova et al. [32] focused mainly on detailed analysis of network structural properties for Bitcoin transaction network. They constructed an aggregated static network for unique users and transactions among them. Then, they evaluated the evolution of network properties over time as well as investigation of major mechanisms to reveal the underlying mechanism that brings about a power-low degree distribution of transaction network. Javarone and Wright [33] conducted a research on two types of networks; the Bitcoin network and the Bitcoin Cash network and analyzed their structures by focusing on global properties including Degree distribution, Clustering Coefficient, and Path Length. Ferratti and D`Angelo [23] considered accounts on the Ethereum blockchain as nodes and the interactions among them as links in the complex network model. They analyzed different measures and metrics at different network snapshots. Shi et al. [17] investigated the anomaly in five leading Bitcoin platforms and by analyzing the normalized logarithmic price return found that the abnormal ask price and bid price appear in bitFlyer at the same time. Sabri-Laghaie et al. [5] suggested a procedure to monitor latent variables of Bitcoin transactions for a binary Blockchain network. They used a hidden Markov multi-linear tensor model (HMTM) and a $T^2$ control chart as a monitoring method for the binary network of cryptocurrency-based transactions. Chen et al. [34]investigated how to monitor and detect structural breaks in dynamic networks. To be more specific, their focus was on identifying anomalies immediately after they occurred. To do so, they applied a Network Sequential Monitoring (NSM) algorithm to the process of network surveillance and used false alarm rates as well as the length of the delay to evaluate the performance of their proposed method. They also used a financial social media dataset form StockTwits, as a crytocurrency network, to indicate the applicability of their method. Based on closing prices and volume datasets, Uras et al.[35]compared linear regression models, as statistical methods, and Artificial Neural Networks, as machine-learning methods, to predict cryptocurrency prices

simultaneously. To obtain accurate price predictions, they employed a novel approach that included short-term regimes. Both machine learning and statistical methods were found to be effective in this study. The structure of crytocurrency networks can be used in other areas such as in storing the data of healthcare systems in a secure and useful manner. For instance, Malathi et al. [36] proposed a blockchain-based management system to indicate the usefulness of such systems in healthcare. A comprehensive review of previous studies in predicting cryptocurrency prices with statistical and machine learning methods was conducted by Khedr et al. [37]. Additionally, they discussed the challenges of forecasting prices using traditional statistical methods and potential research topics.

Machine learning and statistical process control (SPC) methods have shown their applicability and effectiveness in anomaly detection purposes. Among the SPC tools, control charts are by all means the most widely used tools for detecting changes in processes. Despite their applicability for anomaly detection, these techniques have not gained much attention in the monitoring of cryptocurrency transaction networks. To do so, an approach which is based on the statistical monitoring of the transactions, including a SCAN method and a multivariate exponentially weighted moving average (MEWMA) control chart, is applied in this research. For this purpose, a random dynamic network formed by the transactions to consider every cryptocurrency account as a unique node and the transactions between them as edges. Then, the Pribe's scan method is applied to the network model so as to identify any anomaly. After all, a MEWMA control chart is used to monitor the mean of binary Degree, Betweenness, and Closeness centrality measures. Thus, if any abnormality is raised in the number or amount of transactions as an anomaly occurs, the proposed control chart will be able to alert an out-of-control condition.

The remainder of this paper is organized as follows. In section 2 all notations are described. Section 3 presents the network modeling of the problem. Section 4 provides a monitoring scheme. Section 5 states the simulation procedure to assess the performance of the proposed methodology. In Section 6, simulation results and a case study on Bitcoin transactions are offered. Finally, section 7 provides concluding remarks.

## 2. Nomenclature Section

In this section, notations used for the modeling and monitoring of the problem are described as the following.

| Notation | Description |
|---|---|
| $i$, $j$ | Index of nodes |
| $t$ | Index of time period |
| $N$ | Number of nodes |
| $T$ | Number of time periods |
| $N_{IT}$ | Number of simulation iterations |
| $C_D(v)$ | Degree centrality of node $v$ |
| $\sigma_{(u,v)}$ | Number of shortest paths between nodes $u$ and $v$ |
| $\sigma_{(u,v)}(w)$ | Number of shortest paths between nodes $u$ and $v$ that pass through node $w$ |
| $C_B(v)$ | Betweenness centrality of node $v$ |
| $C_B^w(v)$ | Weighted Betweenness centrality of node $v$ |
| $d(u,v)$ | Shortest distance between nodes $u$ and $v$ |
| $C_C(v)$ | Closeness centrality of node $v$ |
| $C_C^w(v)$ | Weighted closeness centrality of node $v$ |
| $O_{t,i}^k$ | Size of the $k$ th neighborhood for the $i$ th node at time $t$ |
| $O_{t,i}^{k^*}$ | The first standardization of $O_{t,i}^k$ |
| $MO_i^k$ | Sample mean of $O_{t,i}^k$ for the $i$ th node over the last $w$ time periods |
| $SDO_i^k$ | Standard deviation of $O_{t,i}^k$ for the $i$ th node over the last $w$ time periods |

| | |
|---|---|
| $M_t^{k^*}$ | Second standardizations of $O_{t,i}^k$ |
| $M_t^k$ | Maximum of $O_{t,1}^{k^*}, O_{t,2}^{k^*}, ..., O_{t,N}^{k^*}$ |
| $MM^k$ | Sample mean of $M_t^k$ at $t$ th time period |
| $SDM^k$ | Standard deviation of $M_t^k$ at $t$ th time period |
| $\lambda$ | Smoothing parameter of the MEWMA statistic |
| $\Sigma$ | Covariance matrix |
| $T_t^2$ | MEWMA statistic for time period $t$ |
| $\alpha$ | Type I error |
| $p$ | Network proportion for applying shifts |
| $\theta_{ijt}$ | Link probability between nodes $i$ and $j$ at period $t$ |
| $w$ | Minimum number of periods for obtaining the SCAN statistic |
| $SP_{SCAN}$ | Signal probability of the SCAN method |
| $SP_{MEWMA}$ | Signal probability of the MEWMA control chart |

## 3. Network Modeling

In this section, some network related definitions are provided then the formulations required for modeling the problem are given. We can define a network [38] as a graph, $G = (N, E)$, which consists of a set of nodes, $N$, and a set of edges, $E$. A graph could be directed or undirected and weighted or unweighted. In this paper, we consider an undirected and unweighted network model, in which nodes are the individuals who buy or sell cryptocurrencies and the interactions among them are considered as edges.

### 3.1. Centrality measures

In graph theory and network analysis, centrality measures represent the most important vertices within a graph. These measures can help identify the most influential nodes in different networks [39]. Various kinds of centrality measures have been developed in the literature of networks. In this research Degree, Betweenness and Closeness centralities are used. Therefore, these centralities are defined in the following.

#### a) Degree centrality

The degree centrality of a node is the number of edges or connections linked to a particular node [40]. It is the simplest centrality measure and can represent the chance of a node in catching the flow (information, money, virus, etc.) of the network. For a given node, $v$, degree centrality can be stated as [41],

$$C_D(v) = \deg(v) \tag{1}$$

In a weighted network, the Degree centrality is the total number of the weight of the edges that are connected to node $v$.

#### b) Betweenness centrality

Betweenness Centrality means how many times a node is in-between and on the shortest path that connects two other nodes [42]. To calculate Betweenness centrality, $\sigma_{(u,v)}$ is considered as the number of shortest paths between nodes $u$ and $v$, and $\sigma_{(u,v)}(w)$ the number of shortest paths between nodes $u$ and $v$ that pass through the node $w$ [43]. The Betweenness of node $w$ is defined as,

$$C_B(w) = \sum_{u,v \in G} \frac{\sigma_{(u,v)}(w)}{\sigma_{(u,v)}} \tag{2}$$

In the weighted Betweenness centrality, $C_B^w(w)$, distances are calculated by the sum of weights on the path between nodes.

#### c) Closeness centrality

Closeness Centrality is calculated as the average of the shortest path length from one node to every other node in the network, which means how close a node is to all other nodes [44]. We consider $d(v,h)$ as the shortest distance between nodes $v$ and $h$. Therefore, closeness centrality is defined as [43],

$$C_C(v) = \sum_{h \in G} \frac{1}{d(v,h)} \tag{3}$$

In the weighted Closeness centrality, $C_C^w(v)$, the

distance is obtained by the sum of weights on the path.

## 4. Monitoring Scheme

### 4.1. The SCAN method

Generally, scan or window statistics are used to find a local signal in spatial data, for example the average pixel value of a picture. Such a local signal is called the local statistic, and the maximum value of such statistic is known as the scan statistic [45]. The scan statistics have a wide range of applications, including epidemiology, reliability and quality control, sociology, telecommunications, geography[46], and even social networks [45, 47]. With this in mind, Priebe et al [45] introduced a moving window-based scan method for graphs, which is based on the maximum of degree, and maximum of neighborhood sizes, to calculate the so-called scan statistics [48-50]. The scan method is used in the current research to monitor the anomalies from a network perspective [48-50].

Suppose that $O_{t,i}^{k}$ is the size of the $k$ th ($k = 0,1,2,...$) neighborhood for the $i$ th node at time $t$. The $k$ th neighbors of a node are all nodes which can be reached from the node in question in exactly $k$ hops. Therefore, the size of the $k$ th neighborhood for a given node is defined as the number of all edges between all the $k$ th neighbors of a node [49]. For instance, in Figure 1, $O_{t,3}^{0}$, $O_{t,3}^{1}$ and $O_{t,3}^{2}$ for the third node at time $t$ are obtained as 5, 6 and 10, respectively.

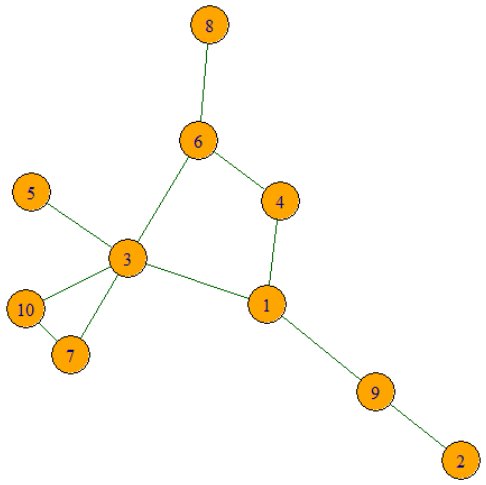



**Fig. 1. A sample network**

To monitor the network changes, statistics $O_{t,i}^{k}$, for $i = 1,2,....,N$ and $k = 0,1,2$, are standardized as $O_{t,i}^{k^*}$ and $M_{t}^{k^*}$. Based on Priebe et al. [45], the first standardization or $O_{t,i}^{k^*}$ for period $t$ is calculated as follows:

$$O_{t,i}^{k^*} = \frac{O_{t,i}^{k} - MO_{i}^{k}}{\max(SDO_{i}^{k}, 1)} \tag{4}$$

Where, $MO_{i}^{k}$ and $SDO_{i}^{k}$ are the sample mean and standard deviation of $O_{t,i}^{k}$ for the $i$ th node over the last $w$ time periods, respectively. Therefore, $MO_{i}^{k}$ and $SDO_{i}^{k}$ are obtained as the following:

$$MO_{i}^{k} = \frac{1}{w}\sum_{j=1}^{w} O_{t-j,i}^{k} \tag{5}$$

$$SDO_{i}^{k} = \left[\frac{1}{w-1}\sum_{j=1}^{w}(O_{t-j,i}^{k} - MO_{i}^{k})^{2}\right]^{\frac{1}{2}} \tag{6}$$

Also, the second standardizations or $M_{t}^{k^*}$, for period $t$, is found as below.

$$M_{t}^{k^*} = \frac{M_{t}^{k} - MM^{k}}{\max(SDM^{k}, 1)} \tag{7}$$

Where $M_{t}^{k}$ is the maximum of $O_{t,1}^{k^*}, O_{t,2}^{k^*}, ..., O_{t,N}^{k^*}$, and $MM^{k}$ and $SDM^{k}$ are the sample mean and standard deviation of $M_{t}^{k}$

at $t$ th time period, respectively. Therefore, $MM^k$ and $SDM^k$ are obtained as the following:

$$MM^k = \frac{1}{w}\sum_{j=1}^{w} M_{t-j}^k \quad (8)$$

$$SDM^k = \left[\frac{1}{w-1}\sum_{j=1}^{w}(M_{t-j}^k - MM^k)^2\right]^{\frac{1}{2}} \quad (9)$$

To monitor a network, statistics $M_t^{k^*}$ for $k = 0,1,2$ and a given $w$ are compared to an upper control limit (UCL). An out-of-control signal is triggered at time $t$, when at least one $M_t^{k^*}$ for $k = 0,1,2$ exceeds the UCL. In this research, $w$ and UCL are set to 5 and 20, respectively [45].

### 4.2. The MEWMA control chart

The MEWMA control chart is used for the simultaneous monitoring of several variables, $\mathbf{X}_t = (x_{1t}, x_{2t}, ..., x_{mt})$, where $m$ is the number of variables [51]. To obtain the MEWMA statistics, vector $\mathbf{Z}_t$ is defined as equation (10).

$$\mathbf{Z}_t = \lambda \mathbf{X}_t + (1-\lambda)\mathbf{Z}_{t-1} \quad (10)$$

where, $0 \le \lambda \le 1$ is the smoothing parameter and , $\mathbf{Z}_t = \mathbf{0}$ [52]. The statistic that is monitored by the MEWMA control chart is given as equation (11).

$$T_t^2 = \mathbf{Z}_t \mathbf{Q}_t^{-1} \mathbf{Z}_t' \quad (11)$$

where, the covariance matrix $\mathbf{Q}_t$ is defined as equation (12) [51].

$$\mathbf{Q}_t = \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]\Sigma \quad (12)$$

In which, $\Sigma$ is the covariance matrix for variables, that is usually estimated based on Phase-I data. An out-of-control signal is triggered when the statistic $T_t^2$ exceeds a UCL. For the purpose of designing the MEWMA control chart, parameters $\lambda$ and UCL should be specified. In this regard, a number of established networks are accumulated over a period of time, and the $T^2$ statistics of the networks are computed. After all, based on the $T^2$ statistics and for a given $\lambda$, the UCL is calculated, thereby a type I error $(\alpha)$ is reached. For Phase-II, new snapshots of networks are amassed and their $T^2$ statistics are examined and contrasted with the UCL. An out of control is alarmed if the $T^2$ statistic of any new network goes over the UCL.

## 5. Simulation Procedure

In this part, to assess the performance of the proposed monitoring methods, the simulation procedure, for finding network changes, is provided. In this regard, $T$ in-control successive networks with $N$ nodes are produced in Phase-I to establish the UCL of the MEWMA control chart and obtain the SCAN statistics. To that end, a link between nodes $i$ and $j$ at time $t$ is defined based on a probability $\theta_{ijt}$. It is supposed that $\theta_{ijt}$ follows a given probability distribution function.

To assess the performance of the proposed procedure, the signal probabilities of simulated networks in Phase-II are computed after applying shifts in the parameters of the probability distribution function of $\theta_{ijt}$, for different proportions of the networks. For this purpose, to produce $T$ random networks in Phase-I and to obtain the SCAN statistics and the UCL of the MEWMA control chart for in-control networks, the next steps are followed:

1) Input $T$, $N$, $\lambda$, $w$, and $\alpha$

2) For $t \in \{1, 2, ..., T\}$

2.1) For node pair $(i, j) \in \{1 \le i < j \le N\}$

a) Based on $\theta_{ijt}$, generate links between nodes.

2.2) For $i \in \{1, 2, ..., N\}$

b) Calculate $O_{t,i}^0, O_{t,i}^1, O_{t,i}^2$,

c) Based on equation (4), calculate $O_{t,i}^{k^*}$ for $k = 0,1,2$.

2.3) Obtain $M_t^k = \max(O_{t,1}^{k^*}, O_{t,2}^{k^*}, ..., O_{t,N}^{k^*})$.

2.4) Based on equation (7), calculate $M_t^{k^*}$ for $k = 0,1,2$.

2.5) Calculate Degree, Betweenness, and Closeness centralities using equations (1), (2), and (3) for all nodes.

2.6) For node pair $(i, j) \in \{1 \le i < j \le N\}$, calculate difference between Degree, Betweenness, and Closeness Centralities, and consider them

as, $C_{D,ij} = |C_D(i) - C_D(j)|$ , $C_{B,ij} = |C_B(i) - C_B(j)|$ , and $C_{C,ij} = |C_C(i) - C_C(j)|$.

2.7) Set $\mathbf{X}_t = (\overline{C_D}, \overline{C_B}, \overline{C_C})$, where $\overline{C_D}$, $\overline{C_B}$ and $\overline{C_C}$ are the averages of $C_{D,ij}$, $C_{B,ij}$, $C_{C,ij}$ over all edges, respectively.

3) For a given $\lambda$ and $t \in \{1,2,...,T\}$ find statistic $T_t^2$ based on equation (11).

4) For all $T_t^2$ statistics, find a UCL that the type I error $\alpha$ is satisfied.

For Phase-II, shifts are considered in different proportions of networks for assessing the performance of the proposed monitoring techniques in identifying anomalies. Without loss of generality, equal link probabilities are assumed for all edges. Then, the following steps are used to assess the signal probabilities:

1) Input $N_{IT}$, $T$, $\lambda$, $w$, UCL, and $p$

2) For $i = 1,...,N_{IT}$,

1.1) Set $count_i^M = 0, count_i^S = 0$

1.2) For $t \in \{1,2,...,T\}$

a) Generate a random network, where the links between $p$ percent of the nodes are defined based on the shifted link probabilities.

b) Perform steps 2.2 to 3 of the Phase-I simulation procedure.

c) For the SCAN method, if $\max(M_t^{0^*}, M_t^{1^*}, M_t^{2^*}) > 5$ , set $count_i^S = count_i^S + 1$ and do step 2 for the next $i$, and for the MEWMA control chart, if $T_t^2 > \text{UCL}$ , set $count_i^M = count_i^M + 1$ and do step 2 for the next $i$.

3. Evaluate $SP_{SCAN} = count_i^S / N_{IT}$ and $SP_{MEWMA} = count_i^M / N_{IT}$

$SP_{SCAN}$ and $SP_{MEWMA}$ are the signal probabilities of the SCAN method and the MEWMA control chart, respectively. In other words, signal probability gives the probability of detecting any change in the network.

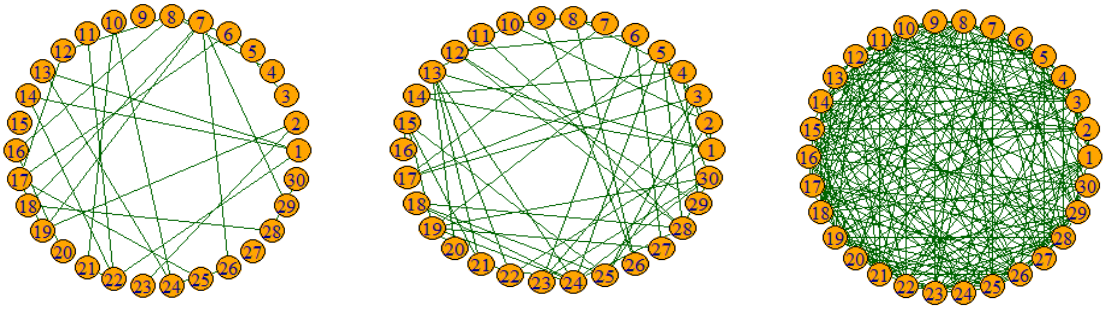
## 6. Experimental Results

In this section, to detect anomalies, the ability of the proposed method is evaluated using the simulation procedure. Then, a real case study on Bitcoin transactions is provided.

### 6.1. Simulation study

The simulation procedure is implemented by setting $N_{IT} = 1000$ , $T = 1000$ , $N =$ 30 and 100, and $\alpha = 0.01$ . Also, it is assumed that $\theta_{ijt} \sim \text{Uniform}(0,0.1)$ , which results in an average probability of $\mu_\theta = 0.05$ . Figure 1(a) represents a sample network with 30 nodes in Phase-I. In Figure 1(b) and Figure 1(c), the network is simulated so that the average probability of connection between 50% and 100% of the nodes is set to $\mu_\theta = 0.5$ , respectively. Similarly, Figures 2 and 3 show the results of the SCAN method and the MEWMA control chart for the proposed shifts. In these cases, 1000 in-control networks are generated, then a shift is applied to the connection probability between 50% and 100% of the nodes. It can be seen that both the SCAN method and the MEWMA control chart are able to detect changes at the start of Phase-II. However, the SCAN method is able to clearly determine the change point of the network.

To evaluate the performance of the SCAN method in Phase-II, 1000 consecutive networks with shifts in %10, %20, %50, %70, and %100 of the networks are generated. In this case, the signal probabilities are obtained for different shifts in the average connection probability between nodes. The simulation results are presented in Tables 1 and 2. As it is noticed, the MEWMA control chart is not able to identify changes when shifts are applied to %10 and %20 of the network. However, the SCAN method is able to identify most of the shifts with high probability. Therefore, the SCAN method outperforms the MEWMA control chart and is efficient in detecting the shifts.



(a) (b) (c)

**Fig. 1. A sample network with (a) no shift, (b) shift in %50 of the nodes, and (c) shift in %100 of the nodes**

### 6.2. The case study

To indicate the usefulness of the proposed procedure, 60 days of Mt. Gox leaked transactions dataset [53] from April to March 2011 is used to monitor Bitcoin transactions over time. Table 3 depicts the structure of the data, with columns representing the information on the source and target nodes, the amount of Bitcoin transacted, the sum of money spent to buy or sell Bitcoin, the money rate, and the transaction time. Each user, a seller or buyer, is defined as a different node, and a connection between every pair of nodes is based on at least one transaction happening between them in a day. Figure 4 depicts the amount of money spent to buy or sell Bitcoin on different days. This can be an indirect indicator of the increase or decrease in the number of transactions.

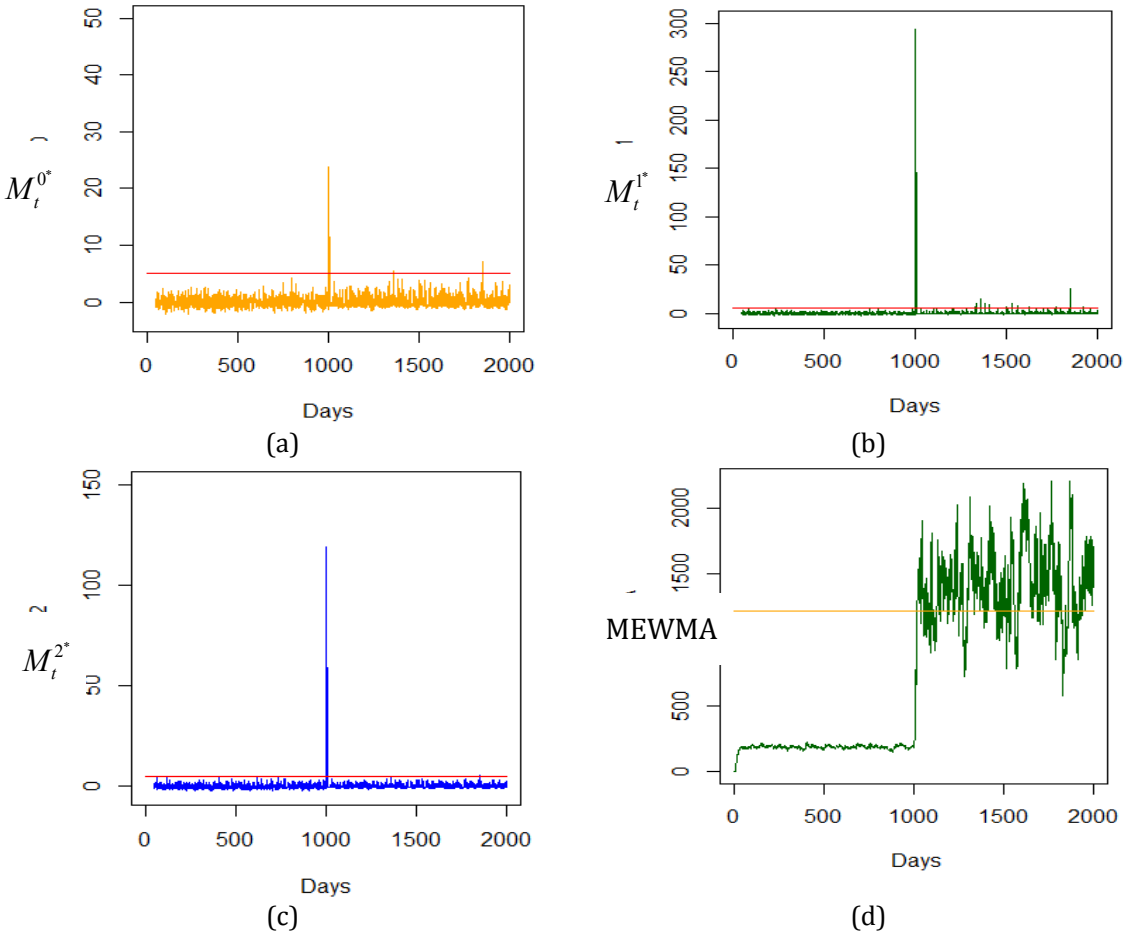



**Fig. 2. (a) $M_t^{0^*}$, (b) $M_t^{1^*}$, (c) $M_t^{2^*}$, and (d) $T^2$ statistics for shift in %50 of the network**

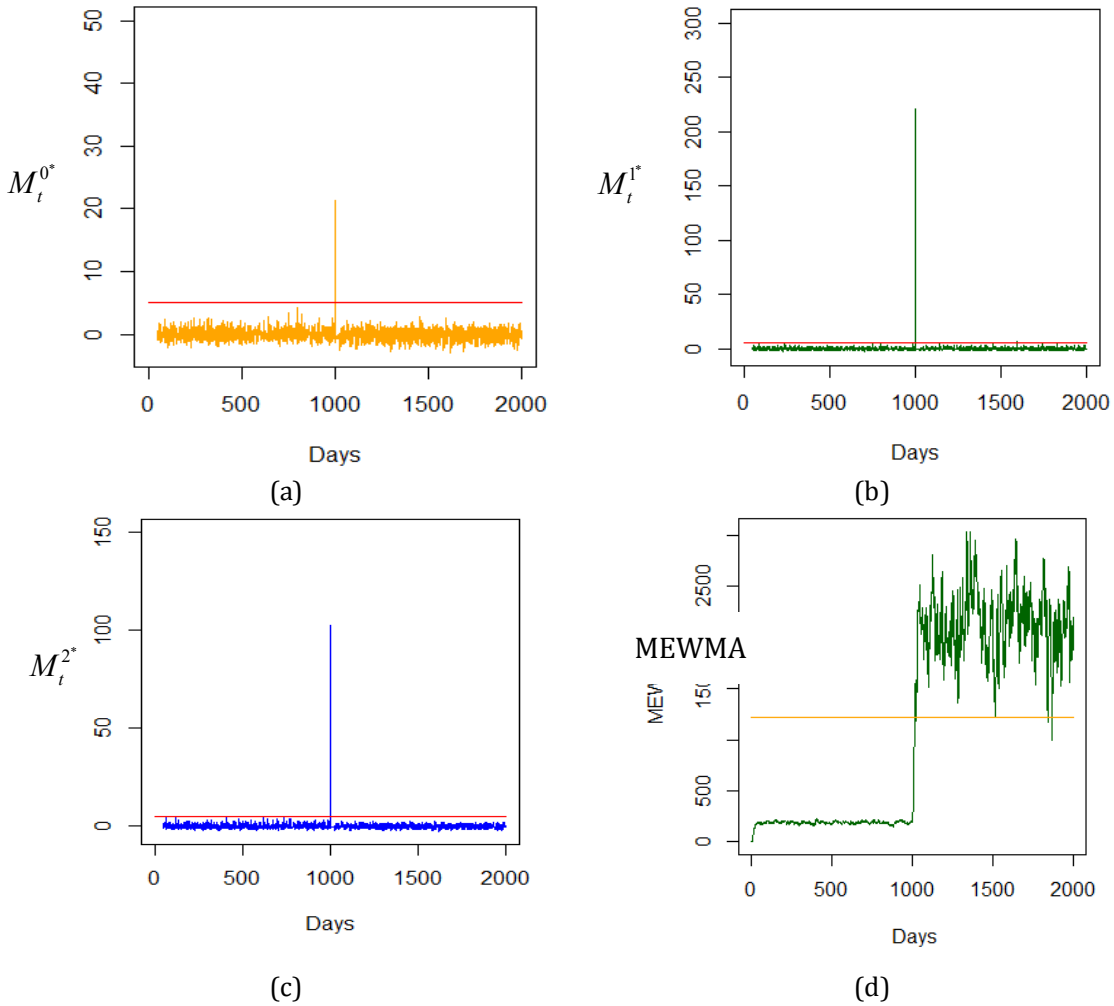



**Fig. 3. (a) $M_t^{0^*}$, (b) $M_t^{1^*}$, (c) $M_t^{2^*}$, and (d) $T^2$ statistics for shift in %100 of the network**

**Tab. 1. The signal probabilities for N=30**

| | Percent | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mu_\theta$ | %10 | | %20 | | %50 | | %70 | | %100 | |
| | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN |
| 0.1 | 0 | 0.045 | 0 | 0.103 | 0.001 | 0.274 | 0.014 | 0.341 | 0.224 | 0.492 |
| 0.15 | 0 | 0.066 | 0.005 | 0.124 | 0.008 | 0.319 | 0.120 | 0.45 | 0.757 | 0.641 |
| 0.25 | 0 | 0.07 | 0.002 | 0.165 | 0.020 | 0.387 | 0.252 | 0.563 | 0.947 | 0.817 |
| 0.35 | 0 | 0.089 | 0.023 | 0.154 | 0.067 | 0.43 | 0.483 | 0.601 | 0.986 | 0.847 |
| 0.5 | 0 | 0.084 | 0.039 | 0.198 | 0.098 | 0.462 | 0.573 | 0.649 | 0.989 | 0.893 |

**Tab. 2. The signal probabilities for N=100**

| | Percent | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mu_\theta$ | %10 | | %20 | | %50 | | %70 | | %100 | |
| | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN | MEWMA | SCAN |
| 0.1 | 0 | 0.971 | 0 | 0.935 | 0 | 0.803 | 0 | 0.499 | 0 | 0.650 |
| 0.15 | 0 | 0.973 | 0 | 0.954 | 0.005 | 0.889 | 0.002 | 0.841 | 0.044 | 0.767 |
| 0.25 | 0 | 0.985 | 0 | 0.982 | 0.010 | 0.443 | 0.021 | 0.887 | 0.219 | 0.856 |
| 0.35 | 0 | 0.988 | 0.001 | 0.985 | 0.022 | 0.437 | 0.006 | 0.618 | 0.609 | 0.902 |
| 0.5 | 0 | 0.991 | 0.031 | 0.987 | 0 | 0.534 | 0.006 | 0.638 | 0.472 | 0.918 |

**Tab. 3. A sample of the dataset**

| Source | Target | Bitcoin | Money | Money Rate | Date |
|---|---|---|---|---|---|
| 895 | 3931 | 23.020 | 18.061 | 0.7845786 | 2011-04-01 00:28:54 |
| 895 | 722 | 10.000 | 7.800 | 0.7800000 | 2011-04-01 00:28:54 |
| 895 | 3605 | 35.000 | 27.300 | 0.7800000 | 2011-04-01 00:28:54 |

Also, Figure 5 illustrates the network of all transactions with 4525 nodes on $T$=1 and $T$=50. It can be seen that there is a notable difference in the transactions on $T$=50. To analyze the network of transactions, the scan method was initially applied to the dataset, which is shown in Figure 6(a). The scan method has detected the signal on $T$=50, which means that there is a rise in the number of transactions starting on $T$=50. Then, by considering the first 49 days as Phase-I and the next 11 days as Phase-II, the MEWMA method was applied. In Figure 6(b), the statistic $T^2$ was plotted over time. It is observed that a rise which began on the 50$^{th}$ day in Phase-II is detected. Comparing Figures 6(a) and 6(b) implies that the signals in both methods are almost concurrent with the increase in the amount of transacted Bitcoins in Figure 5. Therefore, we can conclude that there are significant changes in the daily number and/or amount of transactions.

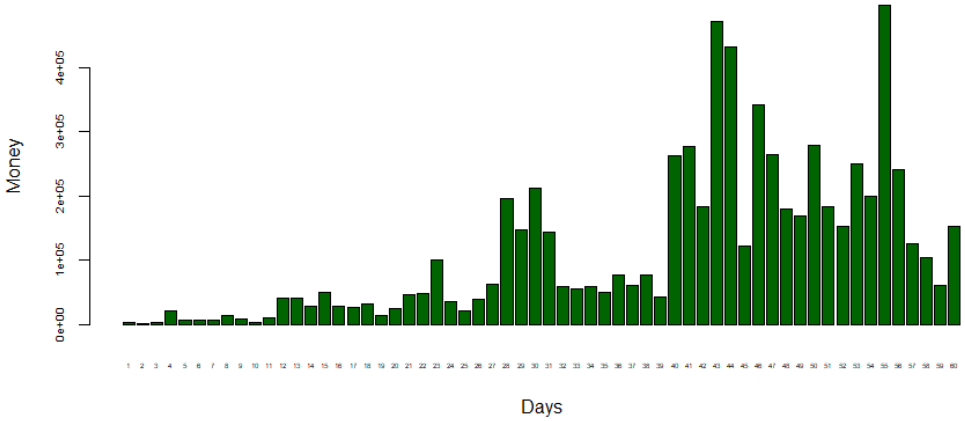
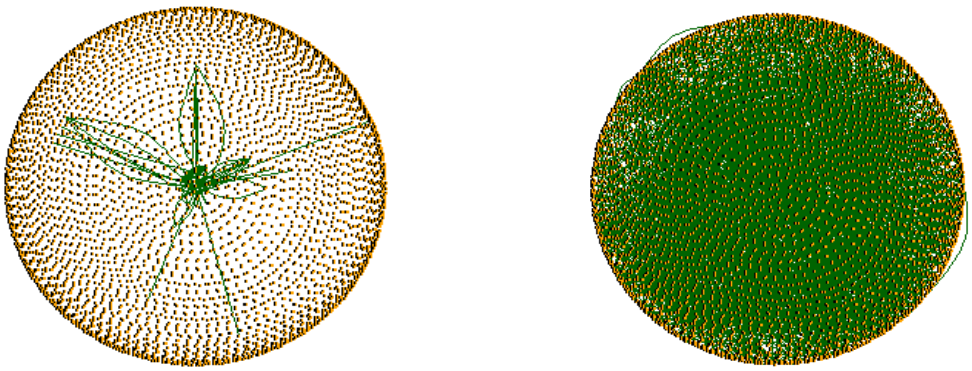
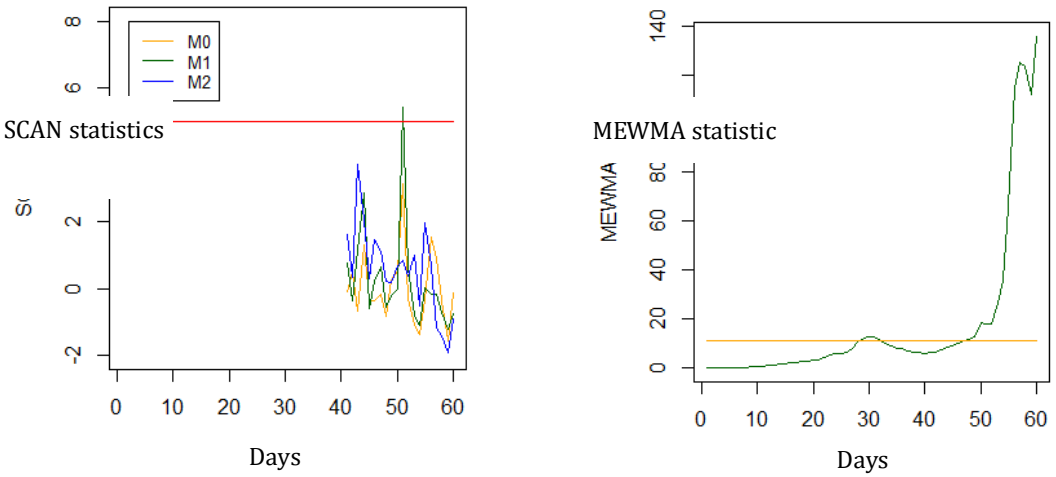



**Fig. 4. The amount of money spent for transactions on different days**



(a) $T=1$ (b) $T=50$

**Fig. 5. All transactions in (a) $T=1$, and (b) $T=50$**




(a) (b)

**Fig. 6. (a) The SCAN method, (b) the MEWMA control chart**

## 7. Conclusion

In this research, a novel routine is proposed to inspect cryptocurrency transactions. To do so, cryptocurrency transactions among blockchain users are modeled as a dynamic network. Then, the size of the first and second neighborhoods is found for the network of transactions. A moving window-based scan method is applied to monitor the network changes. Simulation studies indicate that the SCAN method has the ability to distinguish any changes. As shown in Tables 1 and 2, the MEWMA control chart is unable to identify changes when shifts are applied to the small parts of the networks. The SCAN method, however, is capable of identifying most of the shifts. As a result, the SCAN method outperforms the MEWMA in detecting shifts in the simulation study and is more efficient than the MEWMA. A Bitcoin transactions dataset is applied to represent the usefulness of such a procedure. The results in Figures 6(a) and 6(b) represent that the proposed method detects the signals occurred due to the increase in the number of transacted Bitcoins.

## References

[1] Ahmed, S., *Chapter Three Blockchian and Industry 4.0 Sabbir Ahmed and Razib Hayat Khan.* Blockchain in Data Analytics, (2020), p. 52.

[2] Xu, L.D., E.L. Xu, and L. Li, *Industry 4.0: state of the art and future trends.* International Journal of Production Research, Vol. 56, No. 8, (2018), pp. 2941-2962.

[3] Kamble, S.S., A. Gunasekaran, and S.A. Gawankar, *Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives.* Process Safety and Environmental Protection, Vol. 117, (2018), pp. 408-425.

[4] Bodkhe, U., et al., *Blockchain for industry 4.0: a comprehensive review.* IEEE Access,. Vol. 8, (2020), pp. 79764-79800.

[5] Sabri-Laghaie, K., et al., *Monitoring Blockchain Cryptocurrency Transactions to Improve the Trustworthiness of the Fourth Industrial Revolution (Industry 4.0).* Algorithms, Vol. 13, No. 12, (2020), p. 312.

[6] Kotha, D. and V. Mnssvkr Gupta, *BlockChain: Properties, Application and Bit-coin Case study.* International Journal of Industrial Engineering & Production Research, Vol. 31, No. 2, (2020), pp. 309-315.

[7] Akdoğan, D.A., G.Y.S.e. Kurular, and O. Geyik, *Cryptocurrencies and Blockchain in 4th Industrial Revolution Process: Some Public Policy Recommendations.* Globalisation & Public Policy, (2019), p. 79.

[8] Paliwal, V., S. Chandra, and S. Sharma, *Blockchain Technology for Sustainable Supply Chain Management: A Systematic Literature Review and a Classification*

*Framework*. Sustainability, Vol. 12, No. 18, (2020), p. 7638.

[9] Wu, J., et al., *Analysis of cryptocurrency transactions from a network perspective: An overview*. Journal of Network and Computer Applications, (2021), p. 103139.

[10] Swan, M., *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc." (2015).

[11] Nakamoto, S., *Bitcoin: A peer-to-peer electronic cash system Bitcoin: A Peer-to-Peer Electronic Cash System*. Bitcoin. org. Disponible en https://bitcoin. org/en/bitcoin-paper, (2009).

[12] Nofer, M., et al., *Blockchain*. Business & Information Systems Engineering,. Vol. 59, No. 3, (2017), pp. 183-187.

[13] Mushtaq, A. and I.U. Haq. *Implications of blockchain in industry 4. o*. in *2019 International Conference on Engineering and Emerging Technologies (ICEET)*. (2019). IEEE.

[14] Yli-Huumo, J., et al., *Where is current research on blockchain technology?—a systematic review*. PloS one, Vol. 11, No. 10, (2016), p. e0163477.

[15] Azarnoush, B., et al., *Monitoring temporal homogeneity in attributed network streams*. Journal of Quality Technology, Vol. 48, No. 1, (2016), pp. 28-43.

[16] Mazrae Farahani, E., et al., *Modeling and Monitoring Social Ntwork in term of Longitudinal Data*. International Journal of Industrial Engineering & Production Research, Vol. 29, No. 3, (2018), pp. 247-259.

[17] Shi, F.-B., et al., *Anomaly detection in Bitcoin market via price return analysis*. PloS one, Vol. 14, No. 6, (2019), p. e0218341.

[18] Reid, F. and M. Harrigan, *An analysis of anonymity in the bitcoin system*, in *Security and privacy in social networks*. (2013), pp. 197-223.

[19] Chen, T., et al., *Understanding ethereum via graph analysis*. ACM Transactions on Internet Technology (TOIT), Vol. 20, No. 2, (2020), pp. 1-32.

[20] Zhao, Y., et al. *Exploring eosio via graph characterization*. in *International Conference on Blockchain and Trustworthy Systems*. (2020).

[21] Lin, D., et al., *Modeling and understanding ethereum transaction records via a complex network approach*. IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 67, No. 11, (2020), pp. 2737-2741.

[22] Chen, W., et al., *Dependence structure between bitcoin price and its influence factors*. International Journal of Computational Science and Engineering, Vol. 21, No. 3, (2020), pp. 334-345.

[23] Ferretti, S. and G. D'Angelo, *On the ethereum blockchain structure: A complex networks theory perspective*. Concurrency and Computation: Practice and Experience, Vol. 32, No. 12, (2020), p. e5493.

[24] Alqassem, I., I. Rahwan, and D. Svetinovic, *The anti-social system properties: Bitcoin network data analysis*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 50, No. 1, (2018), pp. 21-31.

[25] Victor, F. *Address clustering heuristics for Ethereum*. in *International Conference on Financial Cryptography and Data Security*. (2020).

[26] Wu, J., et al., *Detecting mixing services via mining bitcoin transaction network with hybrid motifs*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, (2021).

[27] Lin, D., et al., *T-edge: Temporal weighted multidigraph embedding for ethereum transaction network analysis*. Frontiers in Physics, Vol. 8, (2020), p. 204.

[28] Liu, X.F., et al., *Knowledge Discovery in Cryptocurrency Transactions: A Survey*. arXiv preprint arXiv:2010.01031, (2020).

[29] Anoaica, A. and H. Levard. *Quantitative description of internal activity on the ethereum public blockchain*. in *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. (2018). IEEE.

[30] Vidal-Tomás, D., *(In) Stability of the Cryptocurrency Market during the COVID-19 Pandemic: a Network Analysis*. Available at SSRN 3693960, (2020).

[31] Bianconi, G. and M. Agrawal, *Predicting Bitcoin Transactions with Network Analysis*. snap. stanford. edu, (2018).

[32] Aspembitova, A., et al., *Fitness preferential attachment as a driving mechanism in bitcoin transaction network*. PloS one, Vol. 14, No. 8, (2019), p. e0219346.

[33] Javarone, M.A. and C.S. Wright. *From Bitcoin to Bitcoin Cash: a network analysis*. in *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*. (2018).

[34] Chen, C.Y.-H., Y. Okhrin, and T. Wang, *Monitoring network changes in social media*. Journal of Business & Economic Statistics, (2021), (just-accepted): pp. 1-34.

[35] Uras, N., et al., *Forecasting Bitcoin closing price series using linear regression and neural networks models*. PeerJ Computer Science, Vol. 6, (2020), p. e279.

[36] Malathi, M., et al., *Effectiveness of Blockchain Advancement in Patient Statistical Monitoring Network*, in *Data Intelligence and Cognitive Informatics*. (2021), pp. 831-840.

[37] Khedr, A.M., et al., *Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey*. Intelligent Systems in Accounting, Finance and Management, Vol. 28, No. 1, (2021), pp. 3-34.

[38] McGlohon, M., L. Akoglu, and C. Faloutsos, *Statistical properties of social networks*, in *Social network data analytics*. (2011), pp. 17-42.

[39] Newman, M.E.J., *Networks : an introduction*. Oxford; New York: Oxford University Press (2010).

[40] Hansen, D.L., et al., *Social network analysis: measuring, mapping, and modeling collections of connections*. Analyzing social media networks with NodeXL: insights from a connected world. Elsevier Inc, Burlington, (2011), pp. 31-52.

[41] Freeman, L.C., *Centrality in social networks conceptual clarification*. Social networks, Vol. 1, No. 3, (1978), pp. 215-239.

[42] Perez, C. and R. Germon, *Graph Creation and Analysis for Linking Actors: Application to Social Data*, in *Automating Open Source Intelligence*. (2016), pp. 103-129.

[43] Metcalf, L. and W. Casey, *Cybersecurity and applied mathematics*. (2016).

[44] Golbeck, J., *Chapter 3—Network Structure and Measures. Analyzing the Social Web*. Boston: Morgan Kaufmann (2013).

[45] Priebe, C.E., et al., *Scan statistics on enron graphs*. Computational & Mathematical Organization Theory, Vol. 11, No. 3, (2005), pp. 229-247.

[46] Glaz, J. and N. Balakrishnan, *Scan statistics and applications*. Springer Science & Business Media (2012).

[47] Elhambakhsh, F. and M. Saidi-Mehrabad, *Developing a method for modeling and monitoring of dynamic networks using latent variables*. International Journal of Industrial Engineering & Production

Research, Vol. 32, No. 1, (2021), pp. 29-36.

[48] Zhao, M.J., et al., *The effect of temporal aggregation level in social network monitoring*. PloS one, Vol. 13, No. 12, (2018), p. e0209075.

[49] Zhao, M.J., et al., *Performance evaluation of social network anomaly detection using a moving window–based scan method*. Quality and Reliability Engineering International, Vol. 34, No. 8, (2018), pp. 1699-1716.

[50] Kodali, L., et al., *The value of summary statistics for anomaly detection in temporally evolving networks: A performance evaluation study*. Applied Stochastic Models in Business and Industry, (2020).

[51] Crowder, S.V. and S.A. Wiel, *Exponentially Weighted Moving Average (EWMA) Control Chart*. Wiley StatsRef: Statistics Reference Online, (2014).

[52] Lowry, C.A., et al., *A multivariate exponentially weighted moving average control chart*. Technometrics, Vol. 34, No. 1, (1992), pp. 46-53.

[53] Jiajing Wu, J.L., Weili Chen, Huawei Huang, Zibin Zheng, Yan Zhang, *Detecting Mixing Services via Mining Bitcoin Transaction Network with Hybrid Motifs*. arXiv preprint arXiv:2010.01031, (2020).

Follow This Article at The Following Site:

Elhambakhsh F, Sabri-Laghaie K. The SCAN method to monitor cryptocurrency transactions. IJIEPR. 2022; 33 (1) :1-14
URL: http://ijiepr.iust.ac.ir/article-1-1396-en.html