Unsupervised Learning for Robust Bitcoin Fraud Detection

Patrick Monamo*, Vukosi Marivate†, Bheki Twala‡,
*Council for Scientific and Industrial Research
Email: pmonamo@csir.co.za
†Council for Scientific and Industrial Research
Email: vmarivate@csir.co.za
‡University of Johannesburg
Email: btwala@uj.ac.za

***Abstract*—The rampant absorption of Bitcoin as a cryptographic currency, along with rising cybercrime activities, warrants utilization of anomaly detection to identify potential fraud. Anomaly detection plays a pivotal role in data mining since most outlying points contain crucial information for further investigation. In the financial world which the Bitcoin network is part of by default, anomaly detection amounts to fraud detection. This paper investigates the use of trimmed k-means, that is capable of simultaneous clustering of objects and fraud detection in a multivariate setup, to detect fraudulent activity in Bitcoin transactions. The proposed approach detects more fraudulent transactions than similar studies or reports on the same dataset.**

***Keywords*—cybercrime, anomaly, outlier, trimmed k-means, data mining.**

I. Introduction

The latest technological advancement in the global financial system is the establishment of an interned-based payment system capable generation and minting of currency without the use financial institutions as trusted third parties responsible for transaction processing known as the Bitcoin peer-to-peer network [16]. Based on the fact that Bitcoin is recent and involves money or money-equivalents, scepticism relating to usage expansion and its proneness to fraud have been a matter of concern. This research study is commissioned to detect anomalous transactions based on pattern recognition on the Bitcoin network

Bitcoin offers several advantages compared to conventional currency. It eliminates the third party and thus helps lower the transaction fees. This has paved a way for governments across the world to explore the use of Bitcoins for remittance purposes, given the amount of remittance flows that exceed foreign direct investments in most developing countries [14]. Although other currently used technologies for remittance flows managed to lower costs, less than those of conventional banking, a report by the World Bank [22] and G20 [11] shows that the costs are still exorbitant for both the sender and the receiver. The global average cost of sending money is about 8% with the figure escalating to 12% in Sub-Saharan Africa [14], which is still below the global target of 5% thus; the call to explore alternative technological development such as Bitcoin.

Bitcoin is also open to anyone across the globe with no arbitrary legal fees. One more property of Bitcoin, that distinguishes it from other digital currencies and payment systems, is the fact that parties can make payments and transfers anywhere in the world without divulging their true identity [6]. The network makes use of pseudonyms which are addresses derived from public keys. The provision of pseudonyms by the Bitcoin network can be considered to pave a way for cybercriminals to conceal the nature of location, source, ownership, or control of these financial proceeds [15] [5]. This defeats recommendations made by Financial Action Task Force (FATF), which is responsible for standards and promotion of implementation of regulatory and operational measures to combat money laundering and related threads to the global financial system. One of the key requirements deemed to be important towards combating money laundering and terrorism financing relate to identification, verification and reporting. The anonymity of the Bitcoin network bypasses the requirements of FATF according to [22], [5], [14], and hence Silk Road scheme was able to launder approximately $1.2 billion.

Some of the major businesses like Mt Gox and Bitcoinica suffered a loss due to weaknesses attributable to a compromise of one key. To mitigate such weaknesses, latest technological developments that include Hierarchical Deterministic Multisignature (HDM) established from Bitcoin Improvement Proposal such as BIP32 help enhance financial security to an extent that compromising a single party cannot be equivalent to a compromise of funds of other users involved in the sytem [17] [3].

In the advent of a steady absorption of Bitcoin by developing economies for remittances flows, there exist a need for research to develop techniques that will assist regulatory authorities and related law enforcement entities in the fight against cybercrime. The main objective of this paper is to find and classify anomalies on the Bitcoin network based on transaction patterns. This will serve as an aide to detect financial fraud and associated activities such as money laundering. Secondary to that, the paper also seeks to assess performance of the anomaly detection algorithms using publicly available Bitcoin transaction data from blockchain. Furthermore, the study will make an assessment results in relation to the impact of HDM.

This paper is organized as follows: Section II provides an

978-1-5090-2473-5/16/$31.00 ©2016 IEEE

overview of studies related to unsupervised anomaly detection in general as well as specific to the Bitcoin network. The proposed methodological orientation is detailed in Section III. Section IV is dedicated to analysis of experimental results while in Section we provide conclusion and related discussion.

II. Related Work

Most studies involving anomaly detection adopt data with instances labelled as either fraudulent or legitimate [7], [21], [18], [4]. The labelling assists researchers to train anomaly detection algorithms, and assess algorithmic performance with test data. Due to novelty of the Bitcoin network, only a limited number of transactions have been reported as fraud[1] and as such makes the use of supervised technique infeasible.

On the premise of detecting both rogue users and their associated transactions, [19] used k-means clustering, Mahalanobis distance and unsupervised Support Vector Machines (SVM) on 100,000 data points of the Bitcoin dataset due to lack of computational power. [19] employed two types of graphs to model behavioural patterns in the network. Their study considered users as nodes with transactions between them serving as edges and $vice\ versa$. The algorithms used were able to detect 3 out of 30 known cases. In their subsequent study, using the full dataset that was extracted from the network on the 7^{th} of April 2013, [20] attained similar results through the adoption of k-means clustering, Local Outlier Factor(LOF) as well as laws of power degree and densification to attain similar objectives. According to [24], based on Bitcoin network transaction from the genesis block until 13^{th} July 2013, k-means clustering detected two interesting clusters. On the one hand a large cluster contained all good users as well as the known victims while on the other hand the three rogue users were clustered together in the smaller group. Furthermore, [24] generated synthetic node data that resembled the patterns of the three heists under investigation. The performance of the model based on synthetic data was found to attain an accuracy level of 76.5 percent in terms of detection rate. The improvement on findings can be attributable to the not yet uncovered properties of rogue users of the Bitcoin network given the disparities with the real-world data.

K-means clustering has the ability to group instances together, but lacks the prowess of detecting outliers. While LOF is popular for outlier detection, it does not scale well in large datasets with computational time. This paper proposes an approach that will compensate for the above-mentioned weaknesses.

III. Methodology

In this section we provide a brief outline of the dataset used, followed by a description of all features that were extracted from the dataset. The section is concluded by describing the machine learning algorithm proposed for the study. For the anomaly detection techniques, it is assumed that the majority of the transactions on the network are legitimate with at most only 1% being fraudulent.

A. Data Description

This study will use the Bitcoin dataset housed by the Laboratory for Computational Biology at the University of Illinois[2]. All transactions from the genesis block to blockchain 230686 dated 7 April 2013. The blockchain under study contains 6 336 769 users which in our case we refer to as nodes. In between the users are 37 450 461 edges (transactions) that link interactions among users.

B. Feature Extraction

Based on the measured variable provided by the Bitcoin network, we attempt to build more meaningful features that will assist our learning algorithm in terms attaining the desired objectives. A total of 14 features were derived from the transaction data of the Bitcoin network. The following 14 features were derived from the dataset:

- Currency features: total amount sent, total amount received, average amount sent, average amount received, standard deviation received, standard deviation sent
- Network Features: in degree, out degree, clustering coefficient, number of triangles,
- Average neighbourhood (source target) whereby with reference to each query node: source refers to origin on incoming transaction and target is the destination. The four features identified: in-in, in-out, out-out, out-in.

C. Pre−Processing

Given that the dataset lists all transactions that took place during the period under study, cognisance of cases whereby some nodes were involved only in sending or receiving is noted. This led to the existence of missing values in our final dataset and hence imputation was in this regard exercised. We replaced missing values with zeroes based on the premise of equivalence to sending or receiving 0 BTC.

To have appropriate metrics between instances in our multivariate environment, we opted to transform our data. Our transformation resulted in each instance centred around mean zero and unit variance.

D. Proposed Method

The objects contained by the Bitcoin network dataset as described in Section 3.1 are unlabelled, hence this study opt for algorithms that are capable of outlier detection by considering the underlying structure of groupings existing within the network. The compared algorithms are standard k-means clustering and its robust version by [8]. The methods are discussed in the following subsections.

1) K-Means Clustering: On the basis of limited known number of transactions reported as fraud, we opt for k-means clustering. The clustering algorithm comprises of three key steps [23]:

- initialization of the centroids, followed by

[1] https://bitcointalk.org/index.php?topic=576337

[2] http://compbio.cs.uic.edu/data/bitcoin/

- segmenting the dataset into k groups, and
- update the centroids until convergence is attained after several number of iterations

Although [19] cautioned that k-means clustering is not a technique for outlier detection, it lays the basis to evaluate methods given that outliers will be found furthest from the centroids of clusters they are associated with. In k-means, the average behaviour of objects in each cluster is represented by the calculated centroids while the Euclidean distance of each object in a cluster provide us with a measure of location relative to the centre. In this manner the method paves a way towards optimum outlier detection in any given cluster. Based on proposed features extracted, the algorithm is adopted to further confirm evidence provided by previous literature with regard to the potential number of clusters existing within the network.

2) Trimmed K-Means Clustering: The second algorithm is based on partial trimming that is more robust than classical k-means clustering in [8]. Given the trimming level α with the lowest possible variation penalized by Φ, the procedure if formulated as follows:

Let $\alpha \in (0,1)$, the number of clusters k, and the penalty function Φ be given. For any set A such that $P(A) \geq 1-\alpha$ and any k-set $M = m_1, m_2, ..., m_k$ in $\Re^d$, the method considers the variation of M given A to be :

$$V_\Phi^A(M) = \frac{1}{P(A)} \int_A \Phi(\inf_{i=1,...,k} ||X - m||) dP$$

- Obtain k-variations given A, $V_{k,\Phi}^A$ by minimising in M:

$$V_{k,\Phi}^A = \inf_{M \subset \Re^d, |M|=k} \Phi^A(M)$$

- Obtain the trimmed k-variation, $V_{k,\Phi,\alpha}$ by minimizing in A:

$$V_{k,\Phi,\alpha} = V_{k,\Phi,\alpha}(X) = V_{k,\Phi,\alpha}(P_X)$$

$$= (inf_{A \in \beta^d, P(A) \geq 1-\alpha} V_{k,\Phi})$$

The primary objective of the algorithm is to obtain a trimmed set A_0, if it exists, and a k-set $M_0 = m_1^0, m_2^0, ..., m_k^0$, if it exists, through the condition:

$$V_\Phi^A(M_0) = V_{k,\Phi,\alpha}.$$

While standard k-means clustering provides only the properties of the resultant clusters, this robust version assumes the existence of specific proportion of outlier within the network. The technique trims out those objects that are furthest from the centroids to be anomalous in nature, hence the name. According to [12], the general concept of k-means combined with impartial trimming provides robust result in terms of influence function, breakdown point and qualitative robustness as the key performance measures.

IV. Results

This paper used R implementations that have been developed for the above algorithms[3]. The full dataset was aggregated to a node level and the algorithms applied to the first 1 000 000 nodes when instances are .listed according to the increasing node number as per the original transaction data. It should be noted that for model building, all the extracted features were used for both classical k-means and trimmed k-means clustering algorithms. The rest of this sections discusses the results from our outlier detection approaches.

A. Classical k-means clustering

In the absence of knowledge regarding nature of clustering structure of the Bitcoin network, the first step was to estimate the number of clusters k, by using the within sum of squares as the key clustering performance metric. The choice was motivated by its popularity in clustering problems as well as the ease of interpretation. As a result, the clustering algorithm was iterated over a range of values of k to determine the best number of clusters. Given the nature of initial randomization attributable to k-means, the algorithms were ran five times to gauge the stability of the value of k in this regard as well as the centres. Whilst k can be infinitely large, it was only restricted to a maximum of 15 in this particular study to reduce time complexity.

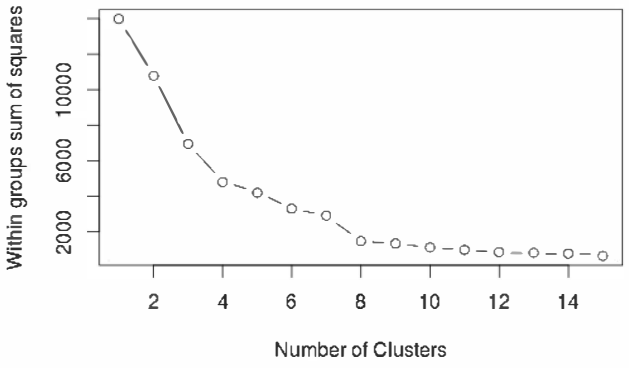

Fig. 1. The elbow chart showing optimal clustering attained at k=8

Figure 1 shows that the optimal number of clusters is realized at k=8. Although k=4 appears to be a good choice, there is a large gain in variability up to k=8 which flattens thereafter, hence, the choice of 8 clusters, which is in agreement with previous reports [19].

In Figure 2 cluster distribution to visualize relative frequencies is provided, while Figure 3 reflects a 2-dimensional plot all resultant clusters as represented by various colour codes for k-means using the two largest components. The distribution shows that we have cluster 8 being the largest with almost 60% of all instances. Due to the algorithm's sensitivity to outliers, this results shows two outright extreme points belonging to a single cluster. Table IV-A below provides a

[3]The main packages used include $mclust$, fpc and $tclust$ [10], [9]

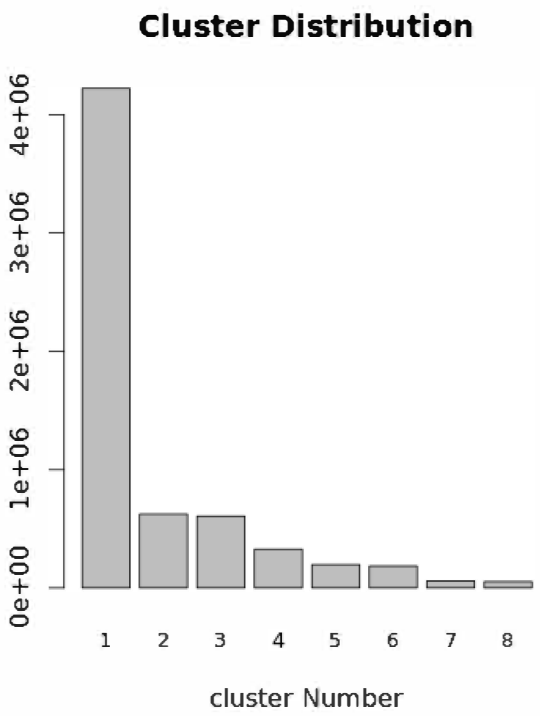

Fig. 2. Distribution of of attained clusters through k-means clustering algorithm.

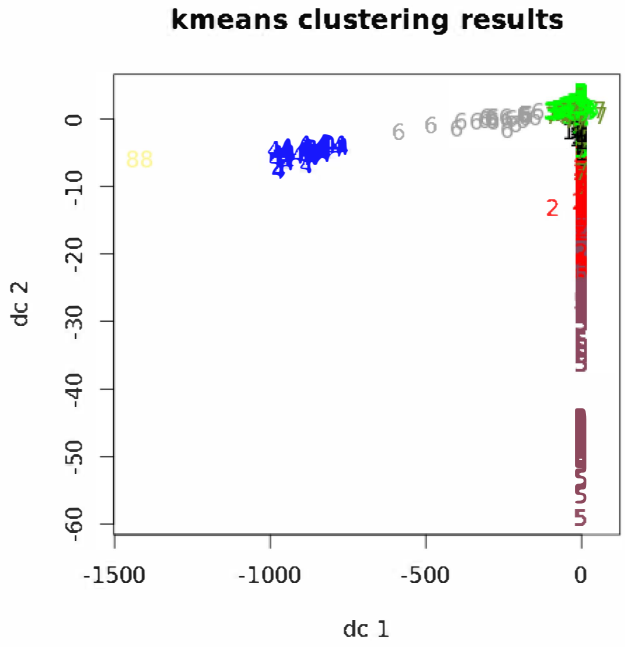

Fig. 3. A graphical representation of k-means clustering results

snapshot of centroids associated with the each of the 8 clusters. The table shows average value of Bitcoin received and sent by each user together with associated user out-degree and clustering coefficient. The number $out-degree$ describes the total quantity of outgoing transactions for each user. This figure appeared fairly moderate among the first 6 clusters followed by and abrupt difference when considering cluster both cluster 7 and 8. The proportion of users associated with each query node that transact together $(clustering-coefficient)$ tend to be strong in the top 3 clusters and poor in the bottom ones. The summary shows that clusters with higher clustering coefficients tend to transact with relatively small amounts of Bitcoins. While this clusters show good connectivity among users as vindicated by higher clustering coefficients, they also exhibit lower values regarding user degrees. In contrast, clusters with poor cluster coefficients were found to have abnormally large node degree and transacting on higher amounts.

TABLE I. CLUSTER CENTROIDS k-MEANS USING SELECTED ATTRIBUTES

ClusterLabel	AverageSent	AverageReceived	ClusterCoeff	OutDegree
1	2.99	2.99	0.50	9.24
2	1.92	1.97	0.61	4.77
3	0.26	0.24	0.70	2.21
4	99.63	107.44	0.23	5.56
5	87.27	64.98	0.31	7.38
6	41.00	36.44	0.12	16.00
7	98.51	67.48	0.00	532 534.00
8	9.50	17.90	0.00	477 035.60

B. Trimmed k-means for robust clustering and outlier detection

Similar to classical k-means, the first steps in this algorithm will be to estimate the number of groups existing withing the network as well as the trimming level, α. Although the actual contamination level is unknown, in this paper it is fixed at $\alpha = 0.01$ as guided by similar studies involving financial fraud and intrusion detection. For comparison purposes with classical k-means, the value of k is restricted to a maximum of 15. The value associated with optimal number of clusters in this method is realized when the difference between the log-likelihood of $k+1$ and k approximates to 0 given a specified trimming level using BIC [13]. The algorithm achieved optimal clustering at k=8 and the distribution with an additional cluster containing outliers is shown on Figure 4

The proportion of objects trimmed as outliers are marked with an "o" on Figure 5.

TABLE II. CLUSTER CENTROIDS TRIMMED k-MEANS USING SELECTED ATTRIBUTES

ClusterLabel	AverageSent	AverageReceived	ClusterCoeff	OutDegree
1	30.01	29.14	0.26	6.10
2	0.04	0.03	0.70	2.00
3	1.70	2.29	0.55	8.19
4	39.84	35.24	0.14	8.63
5	30.51	42.54	0.60	4.92
6	77.04	51.59	0.98	3.81
7	1.02	1.05	0.61	4.51
8	25.86	24.73	0.01	10.45
9	2217.36	1886.70	049	1508.89

1) Linking Results to HDM: Although HDM was not yet implemented at the time of cybercriminal incidents that took place, this research take a further step to assess results in relation to such developments. It is noted that only 5 of the 30

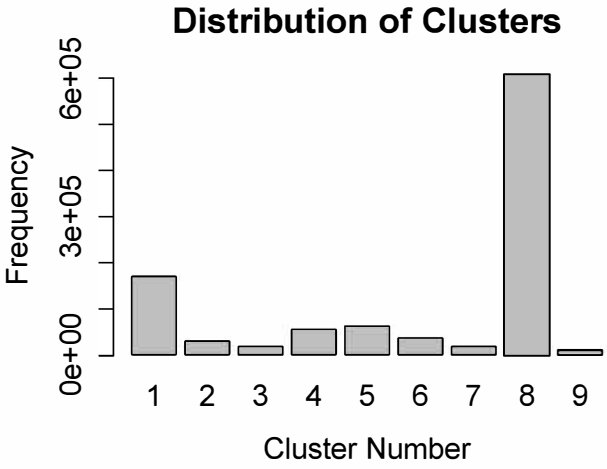

Fig. 4. Distribution of trimmed k-means clustering results

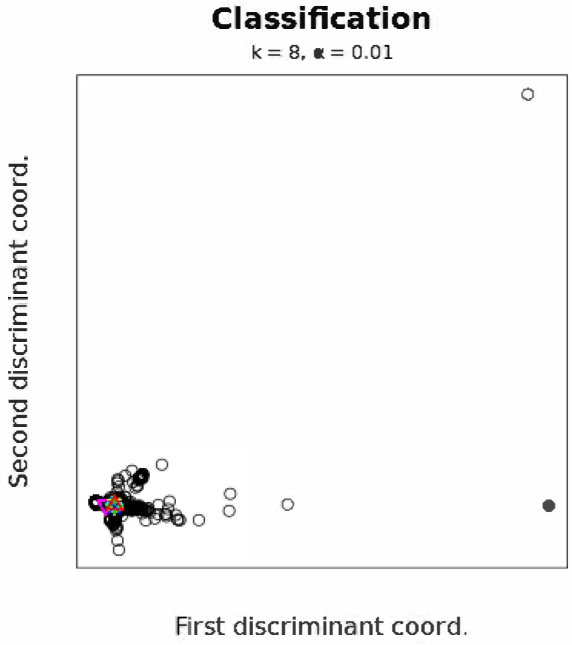

Fig. 5. Clustering results from trimmed k-means with outliers marked 'o'

known anomalies were successfully detected by the algorithm. The detected nodes involved the following entities: 1. Mt Gox, 2. Linode Hack, 3. Stone Man Loss, 4. Allinvain and 5. 50 BTC Theft. Anomalies associated with the first four entities are reported to have originated from poor backups. This fraudulent activities occurred in the absence of HDM which has been developed to mitigates vulnerabilities attributable to wallet backups. The adoption of HDM could have significantly reduced the number criminal incidents experienced by the network users.

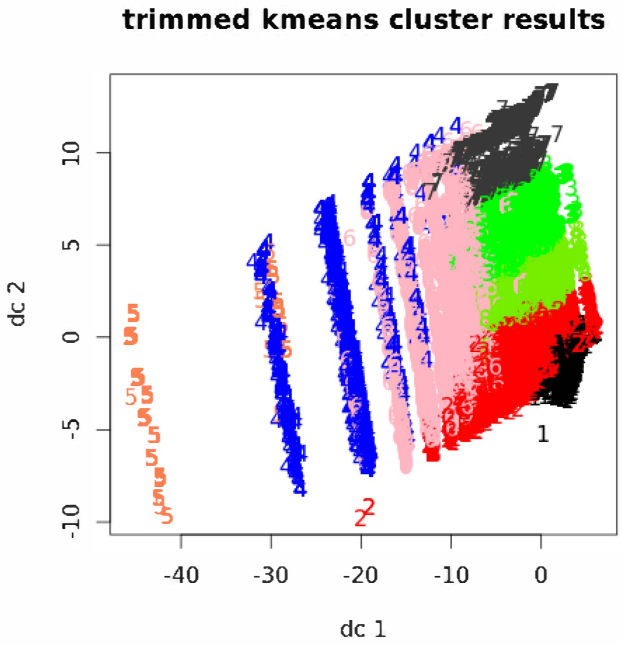

Fig. 6. Clustering results from trimmed k-means with trimmed outlier left out

V. Discussion

The results shows some disparities between the two clustering algorithms in the presence of outliers. K-means due to its sensitivity to anomalies, formed a spurious cluster containing only two objects as depicted in Figure 3. When applying trimmed k-means to the dataset, spurious cluster attained from k-means is filtered out and as a result improvements in group structures is realized. From Table IV-B, it is evident that the detected outliers were spread across multiple clusters as elucidated by moderated attribute values in non-anomalous clusters.The presence of anomalies obstruct visuals of the underlying structure in Figure 5. In Figure 6, all trimmed outliers are left out and an approximately clear structure can be seen.

The detected outliers represented by trimmed proportion was compared with the 30 known fraud cases to assess the performance of the proposed approach. To realize this objective, we use all transactions of the Bitcoin network for the period under study. The raw data of the Bitcoin network contain similar attributes to the list of the 30 known fraudsters. The said attributes in this regard include sender, receiver, date, time (up to seconds level) and value in BTC. Although the list of known criminal elements is made of 30 users, it should be noted that the raw list is composed of 76 transactions which were ultimately matched against the 37 million edges of the network.

Finally, to detect anomalies the two datasets were matched by date, time and amount. Furthermore, the resultant outliers were matched against the outliers detected by the trimmed k-means algorithm. Of the 30 known bale users, the algorithm successfully detected 5 of them.

The findings in this paper proves to be an improvement to results on similar reports in terms of the number known anomalies that were detected successfully. Furthermore, this paper vindicate that the adoption of recent technological developments (e.g. BIP and HDM) when coupled with good performing fraud detection algorithms will enhance financial security of users on the peer-to-peer network. This combination serves as mitigating factors on sceptical attitude towards Bitcoin and thus provide a platform for acceptance by the global village into mainstream economy.

VI. Future Studies

The main challenge in this study that instances are unlabelled and hence becomes difficult to validate results. On the basis of reliance on known criminal elements, future studies will consider comparing results with neighbourhood-based algorithm. In the absences of validation methods in this type of situation, a look at algorithms undersampling majority instances while oversampling majority groups will be explored. Based on the fact that the Bitcoin dataset in in the form of a network, graph-based algorithms for anomaly detection is also an option in this regard.

One of the important challenge in data mining is the ever increasing amount of data which is the case of the transactions in blockchain. From an analytics perspective, it is proposed that techniques on streaming data be considered as well as segmenting the data according to major developments such prior/post HDM.

VII. Conclusion

In this paper, we evaluated the use of trimmed k-means clustering for unsupervised cybercrime detection in the Bitcoin network. Although unavailability of labels which makes it difficult to evaluate algorithmic performance with regard to flagged suspicious activities, the algorithms successfully detected some of the known fraudulent activities. In comparison to previous studies on fraud detection on Bitcoin network, trimmed k-means provided promising results with improvements of detection rate with regard to known fraudulent elements. There is still more work to be done. There is a larger need for advanced feature extraction [1], [2]. With more informative feature extraction, we might be able to train supervised learning methods on the known fraudulent cases and explore if it will reveal other similar behaviour in the network.

References

[1] Leman Akoglu, Mary McGlohon, and Christos Faloutsos. Oddball: Spotting anomalies in weighted graphs. In *Advances in Knowledge Discovery and Data Mining*, pages 410–421. Springer, 2010.

[2] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

[3] Anon. Reclaiming financial privacy with hd wallets, 2013.

[4] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, and J Christopher Westland. Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3):602–613, 2011.

[5] Adrian Blundell-Wignall. The bitcoin question: Currency versus trustless transfer technology. *OECD Working Papers on Finance, Insurance and Private Pensions*, 2014.

[6] Joseph Bonneau, Andrew Miller, Jeremy Clark, Arvind Narayanan, Joshua A Kroll, and Edward W Felten. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Security and Privacy (SP), 2015 IEEE Symposium on*, pages 104–121. IEEE, 2015.

[7] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[8] JA Cuesta-Albertos, Alfonso Gordaliza, Carlos Matrán, et al. Trimmed k-means: An attempt to robustify quantizers. *The Annals of Statistics*, 25(2):553–576, 1997.

[9] Chris Fraley and Adrian E Raftery. Mclust version 3: an r package for normal mixture modeling and model-based clustering. Technical report, DTIC Document, 2006.

[10] Heinrich Fritz, Luis A Garcıa-Escudero, and Agustın Mayo-Iscar. tclust: An r package for a trimming approach to cluster analysis. *Journal of Statistical Software*, 47(12):1–26, 2012.

[11] G20. G20 plan to facilitate remittance flows. Technical report, G20, 2014.

[12] Luis Ángel García-Escudero and Alfonso Gordaliza. Robustness properties of k means and trimmed k means. *Journal of the American Statistical Association*, 94(447):956–969, 1999.

[13] Luis Angel García-Escudero, Alfonso Gordaliza, Carlos Matrán, and Agustín Mayo-Iscar. Exploring the number of groups in robust model-based clustering. *Statistics and Computing*, 21(4):585–599, 2011.

[14] Ralph C Maloumby-Baka, Christian Kingombe, et al. The quest to lower high remittance costs to africa: A brief review of the use of mobile banking and bitcoins. *Centre for Finance and Development Working Paper. Graduate Institute. Geneva*, 2015.

[15] G Krishnapriya MCA and M Prabakaran. An multi-variant relational model for money laundering identification using time series data set. unpublished paper.

[16] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.

[17] Coinbeyond News. Anatomy of hdm structure, July 2014.

[18] EWT Ngai, Yong Hu, YH Wong, Yijun Chen, and Xin Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559–569, 2011.

[19] Phillip Thai Pham and Steven Lee. Anomaly detection in bitcoin network using unsupervised learning methods. Unpublished Report.

[20] Phillip Thai Pham and Steven Lee. Anomaly detection in the bitcoin system-a network perspective. Unpublished Report.

[21] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.

[22] Dilip Ratha, Supriyo De, Ervin Dervisevic, Christian Eigen-Zucchi, Sonia Plaza, and Kirsten Schiettler. Migration and remittances: Recent developments and outlook-special topic: Financing for development. *Migration and Development Brief*, 24, 2015.

[23] Archana Singh, Avantika Yadav, and Ajay Rana. K-means with three different distance metrics. *International Journal of Computer Applications*, 67(10):13–17, 2013.

[24] Deepak Zambre and Ajey Shah. Analysis of bitcoin network dataset for fraud. Unpublished Report, 2013.