International Journal of Digital Crime and Forensics
Volume 13 • Issue 1 • January-February 2021

Detection of Suspicious or Un-Trusted Users in Crypto-Currency Financial Trading Applications

Ruchi Mittal, Netaji Subhas Institute of Technology, New Delhi, India
https://orcid.org/0000-0001-6818-2355

M. P. S. Bhatia, Netaji Subhas Institute of Technology, New Delhi, India

ABSTRACT

In this age, where cryptocurrencies are slowly creeping into the banking services and making a name for them, it is becoming crucially essential to figure out the security concerns when users make transactions. This paper investigates the untrusted users of cryptocurrency transaction services, which are connected using smartphones and computers. However, as technology is increasing, transaction frauds are growing, and there is a need to detect vulnerabilities in systems. A methodology is proposed to identify suspicious users based on their reputation score by collaborating centrality measures and machine learning techniques. The results are validated on two cryptocurrencies network datasets, Bitcoin-OTC, and Bitcoin-Alpha, which contain information of the system formed by the users and the user's trust score. Results found that the proposed approach provides improved and accurate results. Hence, the fusion of machine learning with centrality measures provides a highly robust system and can be adapted to prevent smart devices' financial services.

KEYWORDS

Anomaly Detection, Banking, Centrality Cryptocurrency, Deep Fraud Detection, Financial Services, Machine Learning, Random Forest, Security, Smart Devices, Social Network

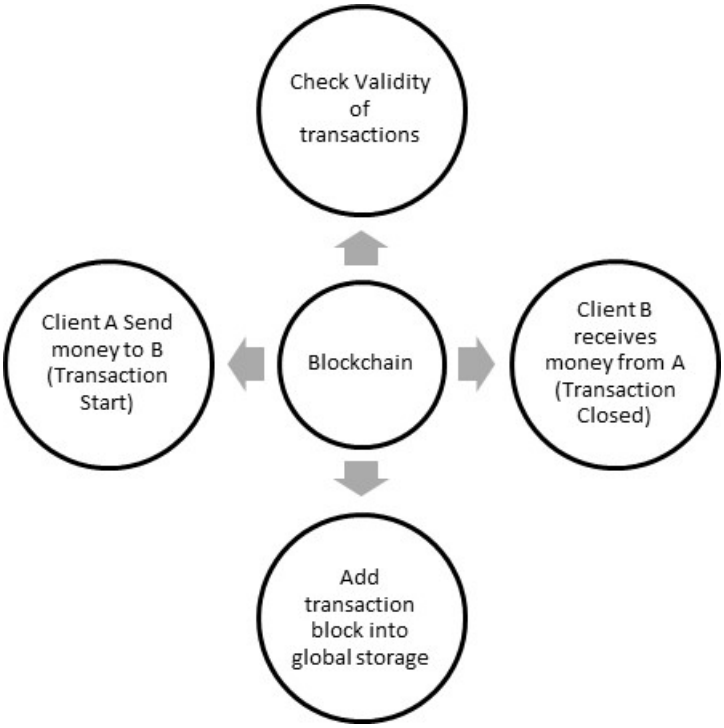
INTRODUCTION

Due to advancements in technologies, online banking services via mobile applications and desktop applications hold continued growth among customers (Teutsch, Jain, & Saxena, 2016). This growth may increase the complexity of the banking system and raise security concerns among users. Also, there are many currencies exhibit in the real world, and each country has its currency, which again increases the complexity of the financial system (Sklavos, & Koufopavlou, 2005). To bypass the banking systems' complexity, a new type of currency came into the picture called cryptocurrency (Ahmad, Kumar, Shrivastava, & Bouhlel, 2018), which has no walls or boundaries and used globally anywhere in the world. The concept of cryptocurrency is new; it may raise security concerns when using mobile applications or desktop applications to prevent fraud and protect personal information.

DOI: 10.4018/IJDCF.2021010105

This article, published as an Open Access article on February 4, 2021 in the gold Open Access journal, International Journal of Digital Crime and Forensics (IJDCF) (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Figure 1. Cryptocurrency workflow



A cryptocurrency is a decentralized form of money compared to any central banking system. In this age, where cryptocurrencies are slowly creeping into the banking services and making a name for them, it is becoming crucially essential to figure out the security concerns when users make transactions. With the high instability of cryptocurrencies and highly competitive central bank currencies, most of the population is hesitant when investing in cryptocurrencies and skeptical of how to differentiate the trusted user, untrusted users. Cryptocurrencies are mainly working with blockchain mechanism, initially proposed by Satoshi Nakamoto in 2009. Satoshi Nakamoto named the proposed bitcoin (Schwartz, Youngs, & Britto, 2014) cryptocurrency. Figure 1 shows the basic workflow of the cryptocurrency using the blockchain mechanism. Here, client A sends money to B without the involvement of any other thirst party.

This paper investigates the untrusted users of cryptocurrency transaction services, which are connected using smartphones and computers. Though the cryptocurrency users are anonymous, it is necessary to maintain the users' reputation information to identify risky users.

As technology is increasing, transaction frauds are growing, and there is a need to detect vulnerabilities in such systems. Several types of risk exist in financial systems, such as hacking, fraud, password leak, and so on (Niu, Ji, & Tan, 2005). Various kinds of attacks can be conducted by attackers to breach smart devices' security and obtain illegal access (Upadhyaya, & Jain, 2016). Such as existing web pages are replaced with fraudulent web pages, broken firewalls for mobile apps using financial services, old desktop apps that are not updated with technologies, and encourage fraud security parameters. Many third parties are involved in financial services that may promote risks, Hackers, or un-trusted users. Hence, authors pick up one of such risks involved in the latest cryptocurrency financial system, i.e., identifying untrusted users from the networks. Such problems motivate us to find chances by collaborating with different domains to find the optimal and reliable

solution. Here, the authors combine social networks, centrality, and machine learning with seeing untrusted users.

The term Social network includes the set of users connected via some mode of relationships. This relationship could be anything such as friendship, business, and so on (Kuncheva & Montana, 2015). In cryptocurrency (Swan, 2015), users are anonymous to each other, and there is direct trading takes place between each other. Based on this relationship, the authors measure the trust factor and find suspicious users in the network.

In-network theories, measuring centralities are among the pre-eminent ways of computing the existence of nodes in the network. Here, the authors utilize centrality measures with machine learning techniques to find suspicious cryptocurrency traders. Centralities are used to find the importance score of individuals in a given network. Many algorithms are defined for measuring the centralities of nodes in the system like degree, eigenvector, stress, page-rank, and so on. Each of these centrality measures has its technique to measure the score of nodes. A degree centrality measure is the amount of all connections of a node. Eigenvector centrality is computed by calculating the relative result of neighboring nodes. Page-rank centrality is given by Google for its search engine to rank the web page.

The machine is one of the leading research areas and could in-cooperate in many applications such as social networks, banking, finance, etc. This happens due to advancements in technology. Such as, People are using machine learning techniques for price predictions in stock markets. Machine learning techniques are supervised and unsupervised — supervised means to develop a predictive model based on both input and output. Unsupervised means grouping or interpretation of data is made using information only. Here, the authors use the supervised learning technique to find the un-trusted users from the cryptocurrency network. They use the methods like SVM classification, linear regression, decision trees, and neural networks in their work. The authors discuss the details of these techniques in the next section.

In this paper, a new algorithm is proposed to identify suspicious users based on their reputation score and collaborate on the concepts of social networks, centrality measures, and Machine learning techniques.

The next section discussed the background and related work in the social network area, machine learning, cryptocurrency, and so on. Part 3 addresses the proposed approach in detail and describes the methodology. Section 4 describes the dataset used for the experiments and the experimental results after applying the proposed approach to the datasets.

BACKGROUND

Cryptocurrency

Due to advancements in Internet usages and technologies, financial services via technologies grow at an exponential rate. Cryptocurrency is one of the emerging financial services, which follows a decentralized mode of payment and services. This decentralization means there is no banking system involved. Cryptocurrency is a decentralized digital financial system, where the encrypted version of currency is generated and distributed in the form (King, & Nadal, 2012). The whole process of cryptocurrency only needed decentralized verification of transactions without any central government. Confirmation of transactions ensures the successful transactions of currency. This process is called mining (Farell, 2015). The whole systems of cryptocurrencies are based on mining, and there are numerous ways in which mining takes place. Cryptocurrencies are the future currencies, which remove all the barriers of centralized systems. Though the overall process of cryptocurrency mining is expensive, yet it provides fast and lightweight services. (Mittal, Arora, & Bhatia2018) present a multi-variant linear regression technique for the price prediction of cryptocurrency.

Nakamoto, in 2008-09 (Nakamoto, 2008), proposes the first decentralized cryptocurrency called Bitcoin. Before the cryptocurrency system, there exist only peer-to-peer currency systems. After bitcoin's success, many types of cryptocurrencies come into the market like etherum (Wood, 2014),

Ripple (Schwartz, Youngs, & Britto, 2014). [24], proposed an anonymous electronic money system, eCash though it works like traditional banking systems. There is much software in the market that users use for transactions without going to the bank, but such software is indirectly linked to banks (Kalodner, Carlsten, Ellenbogen, Bonneau, & Narayanan, 2015). Paypal is such a kind of system, in which users create accounts and bank linked those PayPal accounts with the bank accounts to allow the users to do transactions via the PayPal accounts (Eisenmann, & Barley, 2006). M-Pesa (Jack & Suri, 2011) is again the same kind of system introduced by Vodafone, where the users can add money into their mobile phone account and propel PIN-secured SMS texts to others to do transactions. All these facilities use the centralized currency system, whereas cryptocurrency has its own money (Barber, Boyen, Shi, & Uzun, 2012). The below few steps address the basic working of Cryptocurrencies:

- An account with a unique formal address is assigned to each cryptocurrency user. The legal address is an individual or global identification of a user.
- The payer made the transaction to the payee's address, and cryptograms use the payer's private key.
- The mining process confirms the transaction (Teutsch, Jain, & Saxena, 2016).

Security in Social Network

(Wadhwa & Bhatia, 2015) proposed an approach based on the Markov chain for identifying the radicalization in social networks. (Wadhwa, Bhatia, 2016) proposed new metrics and an improved algorithm for finding the radicalization and show the impact on Twitter datasets. (Wadhwa & Bhatia, 2014) present a survey on various types of radicalization that exist in twitter datasets. (Mittal, Bhatia, 2018) proposed an approach for finding anomalous users in multi-layer networks using the Gaussian distribution technique. (Mittal & Bhatia, 2017) proposed a methodology that identifies structural hole nodes from multiplex networks.

Social Network and Centralities

Modeling of many real-life systems in the form of the social network allows researchers to analyze strategies in many ways, such as the structure of the system, interaction among entities in the network, etc. (Newman, 2004, 2001) proposed many algorithms for community detection, centrality measurements, and so on in social networks. (Battison & Nicosia, 2014) proposed algorithms for finding communities and measuring centralities for multi-layer systems.

(Wadhwa & Bhatia, 2012) the introduction of social networks, the current trends of social networks, and the future implication. (Wadhwa & Bhatia, 2014) Present a survey on community detection algorithms and classify them as per technique used and applications. (Mittal & Bhatia, 2017) present an approach for characterizing the social network based on the community structure. (Mittal & Bhatia, 2018) suggest a study to describe the various properties of social networks such as community, centralities, etc. on multi-layer systems. (Mittal & Bhatia, 2018) find prominent authors from the scientific collaboration network using social network techniques. (Huang, Shao, Wang, Buldyrev, Stanley, & Havlin, 2013) (Shao, Huang, Stanley, & Havlin, 2014) Proposed algorithm for finding clusters in social networks and check the robustness of those clusters with other techniques. Hence, social network analysis and its methods are useful in many ways in various applications.

Here, we want to find suspicious or untrusted users based on the trust score given to each edge of the network, including the concept of centrality measures in their work. There is a direct interaction between users in cryptocurrency transactions, and this interaction forms a system.

Centralities are one of the main ways of finding the importance of nodes in networks. (Girvan & Newman, 2002) Check the behavior of nodes by computing the centrality measures. There are many algorithms for centrality measurements such as degree, Katz, closeness, eigenvector, page rank, etc. (Mittal & Bhatia, 2018) (Mittal & Bhatia, 2019) proposed closeness centrality and bottleneck centrality formulation in multi-layer social networks. (Mittal & Bhatia, 2018) offered to collaborate

nature-inspired algorithms with social networks' structural properties to find the centrality and structural holes in multi-layer systems. (Bavelas,1950) propose the very first formulation of a closeness centrality algorithm, which computes the closeness centrality by adding the shortest distances from a node to every other node. Eigenvector centrality is calculated by observing the relative score of the neighboring nodes (Newman, 2014).

Machine Learning

The area of machine learning grows exponentially in research and is used in many real-life applications. It is of two types: supervised & unsupervised learning. Many researchers have done lots of work on both kinds of techniques (Ghani, Jones, & Rosenberg, 2003) (Nigam, 2001). Classification is from supervised learning techniques, where each data belongs to a class, and there is a fixed set of categories (Mitchell, 1997) (Niu, Ji, & Tan, 2005). Here, we focus on supervised learning, primarily considering the input and output of both. Random forest, Naive Bayes, SVM classification, and Neural networks are the few classification methods (Escudero, Màrquez, & Rigau, 2000) used to predict the values for users in cryptocurrency dataset.

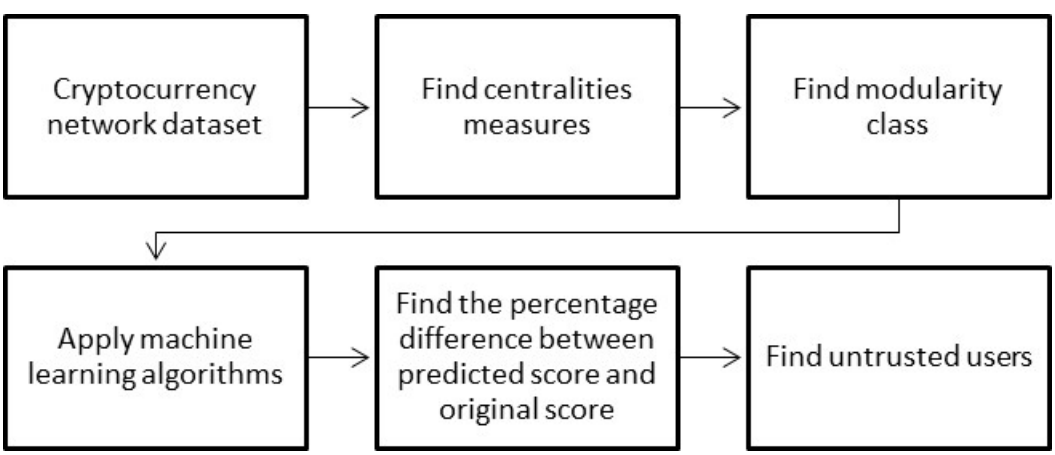
PROPOSED APPROACH

Here, we discuss the proposed approach for finding the untrusted users from the cryptocurrency network in detail. The proposed method starts with computing the system's centralities, such as degree, weighted degree, closeness, and page rank, along with the modularity class. New centrality values are predicted for each node using machine learning algorithms. The modularity class is kept as the class parameter for classification. Finally, the user score is calculated for each node by computing the difference between the original score vs. the predicted score of the nodes. The nodes having a high percentage difference could be untrusted users. The untrusted users are computed by observing the nodes; if a node had a less original score, the predicted number is too low, which is an intimate detail about the approach. Figure 2 shows the workflow of the proposed methodology.

Centrality Measures

- **Degree Centrality:** It is the sum of the count of the incoming and outgoing connections to other nodes (Daly, & Haahr, 2007) (Watanabe, & Kabashima, 2014)). In other words, for a given graph G having node-set A and edge set B is given as:

Figure 2. Workflow of the proposed approach



$$D(a) = \deg(a)$$

The purpose of using degree centrality is that it is commonly used centrality algorithms based on the total links made by a node to other nodes. Here, degree centrality finds the untrusted users. A user with a low degree of centrality is marked as the untrusted user.

- **Weighted Degree Centrality:** It is a one-step forward version of the degree centrality. A new metric 'weight' is also considered for the computation of degree centrality (Tang, Wang, Zhong, & Pan, 2014). The weighted degree centrality is the weighted mean of connection per node in the graph. Here, we are using the weighted network datasets, i.e., each edge of the network holds some weight or the predefined trust score between users of the system. So, they compute this centrality for each of the nodes to ensure the results.
- **Closeness Centrality:** It is based on the shortest path measures, which means closeness centrality finds a node's reachability to all other nodes in the network (Mittal, & Bhatia, 2018). It is computed as the reciprocal of the sum of a node's shortest paths to all other nodes.
- **Page-Rank Centrality:** Google proposed this centrality, and they made this algorithm for predicting the pages for the desired search. It works by counting the number and quality of links to a page to determine a rough estimate of how important the website is (Benincasa et al., 2006).

Modularity Class

The term modularity means the strength of a network's sub-networks, and modularity class means a highly connected group of nodes (Grandjean, 2015).

Machine Learning Techniques

In this paper, the author uses the following machine learning techniques:

- **SVM Classification:** "Support Vector Machine" (SVM) is one of the popular supervised learning algorithms used for classification and regression problems. In this classification mechanism, each data item is plot into an n-dimensional region with the value of each feature being the value of a distinct coordinate (Fei & Liu, 2006). Then, by analyzing the hyperplane, which discriminates the different classes, performs classification. It is a binary classifier that classifies the units into two categories, i.e., valid or invalid.
- **Naive Bayes:** Naive Bayes technique follows a probabilistic approach, which means such methods follow the statistical process to find the set of probabilistic features (Escudero, Màrquez, & Rigau, 2000). It mainly follows the concept of Bayes theorem for finding the healthy norms between components. It is especially beneficial for large datasets, and it is a highly sophisticated classification method.
- **Decision Trees:** It is a prediction model with sense-tagged corpus, where training occurs (Navigli, 2009). It follows yes-no kind of rules for classification and division of the training dataset. The decision tree structure considers internal nodes as features, and edges represent feature value and leaf nodes as sense. The decision tree classification starts from the top to bottom or from the root node to the leaves. Every node of the tree denotes the test instance for some attribute, and every edge declining from that node matches one of the possible answers to the test instance. In the below diagram, we show the working of the decision tree algorithm.
- **Neural Networks:** The concept of neural networks is based on the interconnection of artificial neurons. It includes many hidden layers for the processing of inputs. Input and the expected output are used for learning techniques in neural networks. It follows either a backpropagation-based

Figure 3. Decision Tree Algorithm

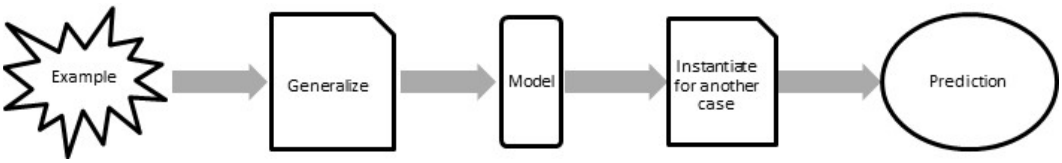
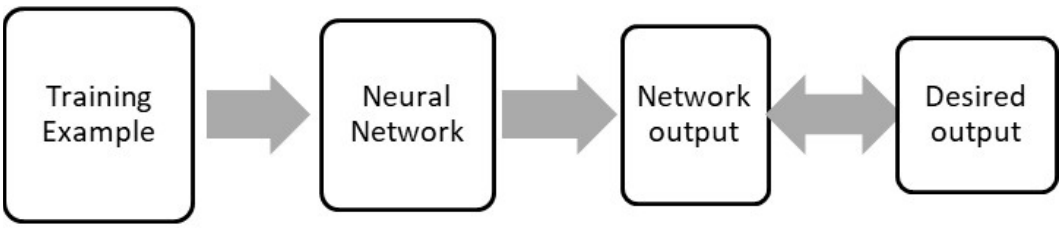


Figure 4. Sample Neural Network



feedforward network. This approach aims to use input features to partition the training set into non-overlapping sets according to expected outputs (Azzini et al., 2008).

Users Score

The user score is computed by taking the difference between the nodes' original centrality score and the predicted score of the nodes. The nodes having high contrast could be marked as untrusted users. Also, if a node had a less original score and the expected number is too low is marked as an untrusted user. The proposed algorithm is given below.

Algorithm 1. Prediction of Untrusted users in cryptocurrency Network

```
Data: Graph G
Result: utu: vector stores set of untrusted users of the network
Initialization:
  1.  Wdcs: weighted degree centrality set
  2.  Cdcs: closeness degree centrality set
  3.  Prdcs: page rank degree centrality set
  4.  Dcs: Degree centrality set
  5.  for each node value ∈ wdcs, cdcs, prdcs, dcs
  6.  Apply SVM classification by considering wdcs, cdcs, prdcs,
      dcs as multiple features
  7.  num_fea= [wdcs, Cdcs, Prdcs, Dcs]
  8.  for i in num_fea:
  9.    find mean and standard deviation
  10. Train the dataset
  11. Apply SVM algorithm
  12. Predict the value for each node individually for each feature
  13. Find diff = predicted_value - actual_value
  14. if diff > Y or diff < A
  15. mark node as untrusted
```

```
16. utu_svm[i] = node_id
17. repeat step 7-41 for naïve bayes, decision tree and neural
    network method
18. utu is the desired output
```

The working of the above algorithm starts by calculating the different centralities for each node using the dataset. Here, the authors make use of the 'Gephi' tool for computation. From steps 1-4, the authors store the value of each node of different centralities. In step 5-6, the authors consider each of the centrality as a feature and apply SVM to it. This algorithm is implemented in the Python programming language. From step 7-12, they define each component, apply SVM using modularity class, and compute the same accuracy score. In step 13, they calculate the difference between the computed value and each node's actual value. From 14-16, check if the difference is less than some threshold or some value marked value, then the node is marked as the untrusted node and put that node into the utu set. Repeat the above steps for other methods.

Experimental Results

Datasets Used

The proposed model is tested against two cryptocurrency dataset- Bitcoin-OTC and Bitcoin-Alpha (Leskovec, Krevl, & Datasets, S. N. A. P. 2014). Both the dataset includes the who trusted whom network of people trading bitcoin cryptocurrency—analyzing the trust forms both the cryptocurrency social network datasets. As we studied earlier, the users of many cryptocurrencies are anonymous to one another, and it is a need to keep track of a record of users' reputation to restrict transactions among untrusted users. The details of the datasets used are described in Table 1.

Performance Metrics

The performance metrics that have been used for detection of untrusted users are done using precision, recall, and f1-score. Precision is the ratio of correctly predicted positive value for the users to the total predicted the positive value for the users. The recall is the ratio of correctly predicted positive value for the users to all value for the users. F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Classification accuracy is an important performance metric, which is given by the percentage of users rating score matched to available users score.

Results

This section covers the discussion on untrusted user detection results along with various performance metrics for the BITCOIN-OTC and BITCOIN-ALPHA dataset. Table 5 and Table 6 show the low ranked users from the BITCOIN-ALPHA and BITCOIN-OTC dataset, having low centrality values, and those nodes could be untrusted users.

Table 1. Bitcoin OTC and Bitcoin ALPHA Dataset

Parameters	Values for Bitcoin-OTC	Values for Bitcoin-Alpha
Nodes	5,881	3,783
Edges	35,592	24,186
Range of edge weight	-10 to +10	-10 to +10
Percentage of positive edges	89%	93%

In table 2 and table 3, the authors show the performance obtained from machine learning algorithms and centrality methods applied to both the datasets. The precision, recall, and f1-score are calculated for each machine-learning algorithm used to many centrality measures. For example, for the bitcoin-alpha dataset, precision-recall value is less for eigenvector centrality in SVM classification, whereas it is high in the Random forest. It is also found that the overall performance of the neural network is inadequate for both the dataset. So, the machine learning algorithm's choice is again an essential aspect of such kind of problems.

Table 2. Precision, Recall, F1-score and Support metrics for Bitcoin-Alpha Dataset

Machine Learning Algorithm	Centrality Method	Precision	Recall	F1-Score	Support
SVM	Degree	0.00	0.00	0.00	9
	Weighted Degree	0.67	0.10	0.17	20
	Closeness	0.80	0.71	0.75	123
	Page Ranks	0.71	0.26	0.38	38
	Eigenvector	0.59	0.66	0.63	131
	Average	0.55	0.35	0.39	321
Random Forest	Degree	0.89	0.89	0.89	9
	Weighted Degree	0.89	0.85	0.87	20
	Closeness	0.98	0.98	0.98	123
	Page Ranks	0.83	0.79	0.81	38
	Eigenvector	0.94	0.99	0.96	131
	Average	0.96	0.90	0.90	321
Neural Network	Degree	0.00	0.00	0.00	9
	Weighted Degree	0.00	0.00	0.00	20
	Closeness	0.00	0.00	0.00	123
	Page Ranks	0.00	0.00	0.00	38
	Eigenvector	0.29	0.99	0.44	131
	Average	0.96	0.90	0.90	321
Naïve Bayes	Degree	1.00	1.00	1.00	9
	Weighted Degree	1.00	1.00	1.00	20
	Closeness	1.00	1.00	1.00	123
	Page Ranks	1.00	1.00	1.00	38
	Eigenvector	1.00	1.00	1.00	131
	Average	0.98	0.98	0.98	321
Decision Tree	Degree	1.00	1.00	1.00	9
	Weighted Degree	1.00	1.00	1.00	20
	Closeness	1.00	1.00	1.00	123
	Page Ranks	1.00	1.00	1.00	38
	Eigenvector	1.00	1.00	1.00	131
	Average	0.97	0.97	0.97	321

Table 3. Precision, Recall, F1-score and Support metrics for Bitcoin-Alpha Dataset

Machine Learning Algorithm	Centrality Method	Precision	Recall	F1-Score	Support
SVM	Degree	0.32	0.97	0.48	295
	Weighted Degree	0.69	0.22	0.33	110
	Closeness	0.00	0.00	0.00	6
	Page Ranks	1.00	0.08	0.15	12
	Eigenvector	0.94	0.33	0.49	52
	Average	0.71	0.47	0.45	1177
Random Forest	Degree	1.00	1.00	1.00	295
	Weighted Degree	0.96	0.98	0.97	110
	Closeness	0.83	0.83	0.83	6
	Page Ranks	0.96	0.94	0.95	12
	Eigenvector	0.00	0.00	0.00	52
	Average	0.87	0.88	0.87	1177
Neural Network	Degree	0.79	0.13	0.22	295
	Weighted Degree	0.20	0.98	0.34	110
	Closeness	0.00	0.00	0.00	6
	Page Ranks	0.00	0.00	0.00	12
	Eigenvector	0.29	0.99	0.44	52
	Average	0.27	0.23	0.13	1177
Naïve Bayes	Degree	0.99	1.00	1.00	295
	Weighted Degree	1.00	0.98	0.99	110
	Closeness	1.00	1.00	1.00	6
	Page Ranks	1.00	0.92	0.96	12
	Eigenvector	0.96	1.00	0.98	52
	Average	0.93	0.94	0.93	1177
Decision Tree	Degree	1.00	1.00	1.00	295
	Weighted Degree	1.00	1.00	1.00	110
	Closeness	1.00	1.00	1.00	6
	Page Ranks	1.00	1.00	1.00	12
	Eigenvector	1.00	1.00	1.00	52
	Average	0.94	0.94	0.94	1177

For both the datasets, the average accuracy for all classification algorithms in terms of accuracy is given in table 4. The accuracy concludes that the Naïve Bayes algorithm's performance is better compared to other classification algorithms for both the datasets. So, we choose the Naïve Bayes algorithm to calculate the users' scores for the next step of the algorithm.

At last, the user score for each node is computed from predicted values to the original values. It is found that more than 30% of nodes in the bitcoin-alpha network have a high difference between the predicted vs. original score, and such nodes are marked as untrusted users. Table 5 shows the top-10 untrusted users of the bitcoin-alpha networks.

Table 4. Accuracy of the proposed model for BITCOIN-ALPHA and BITCOIN-OTC Dataset

Machine Learning Algorithm	Accuracy (Bitcoin-Alpha)	Accuracy (Bitcoin-otc)
SVM	96.5%	96%
Neural Networks (200 hidden Layers)	82.5%	81%
Naive Bayes	96.5%	97.5%
Random Forest	94%	93.5%
Decision Tree	96%	95.5%

Table 5. Top-10 untrusted users in BITCOIN-ALPHA dataset

Node_Id	Degree	Weighted Degree	Closeness Centrality	Modularity Class	Page Ranks
3363	1	1	0	0	0.000072
7348	1	-1	0	1	0.000079
797	1	10	0	2	0.000078
805	1	10	0	2	0.000075
873	1	9	0	2	0.000082
1296	1	5	0	2	0.000071
2449	1	1	0	2	0.000079
2791	1	1	0	2	0.000076
2835	1	1	0	2	0.000078

For bitcoin-otc, the percentage of each node's score from predicted values to the original amounts finds that more than 35.5% of nodes in the bitcoin-otc network have a high difference between the predicted vs. original score. Such nodes are marked as untrusted users. Table 6 shows the top-10 untrusted users of the bitcoin-otc networks.

Table 6. Top-10 untrusted users in BITCOIN-OTC dataset

Node_Id	Degree	Weighted Degree	Closeness Centrality	Modularity Class	Page Rank Centrality
1110	2	6	0	0	0
3363	1	1	0	0	0
7348	1	-1	0	0	1
797	1	10	0	0	2
805	1	10	0	0	2
873	1	9	0	0	2
1296	1	5	0	0	2
1646	4	3	0	0	2
2449	1	1	0	0	2

The proposed algorithm results are compared with some similar methods and found that the performance of the proposed algorithm is better compared to others. First, the comparison is made with the method proposed by (Heard, N. A., Weston, Platanioti, & Hand, 2010). It is found that the number of untrusted users found from the proposed methodology and heard's procedure is quite similar. There is >90% of untrusted users are seen the same from both the methods. Results are also compared with the methods proposed by (Mittal, & Bhatia, 2018) and found that more than 83% of untrusted users are the same from both the approaches.

These results conclude that when we apply the concepts of machine learning and social network techniques together to the cryptocurrency financial systems, we can quickly identify the untrusted users or suspicious users in the network and restrict such users for further frauds in transactions.

Implementation in Cyber Hack Applications

The significant threat of many financial services is hacking. By hack mean breaking the security walls set to prevent the financial data. Many software engineers are involved in cyber hacking called hackers and do illegal work. The cyber hack is one of the most significant hazards in cryptocurrencies transactions, and many cybersecurity companies always keep an eye on such cyber hackers and develop applications to stop such peoples.

The proposed architecture can be used to identify such hackers from all such application, which are involved in financial activities. This step increases the trust among cryptocurrency users and enhances the security walls for such malicious users.

CONCLUSION

This paper proposes a new approach for identifying the untrusted or suspicious users in futures banking and financial services called cryptocurrencies. The proposed method collaborates with the structural properties of social networks (Centralities and modularity class) and machine learning techniques (SVM, Random forest, and so on). The experimental results show that nodes with low weight and a smaller number of connections could be untrusted users. The proposed approach's main advantage helps to identify malicious users from financial treading applications and helps to increase security walls for users who are treading over cryptocurrency fearlessly. Thus, the methodology is vigorous, can be used to secure the crypto-currencies financial trading and usage, and ensures that untrusted users are quickly identified and may block further financial services. Hence, the collaboration of Machine learning techniques enhances the traditional methods for identifying untrusted users and results. The proposed model may extend to deep learning techniques and more real-world applications hookup with cryptocurrencies in smart devices for future work.

ACKNOWLEDGMENT

The authors would like to thank the Snap-Stanford databank repository for providing the Bitcoin-OTC and Bitcoin-ALPHA dataset. These datasets have been used to validate the proposed methodology and predict the un-trusted users involved in bitcoin trading. Authors also like to acknowledge friends, family, and colleagues for their support and effort to complete the work successfully.

REFERENCES

Ahamad, S., Nair, M., & Varghese, B. (2013, May). A survey on crypto currencies. In *4th International Conference on Advances in Computer Science, AETACS* (pp. 42-48). Citeseer.

Ahmad, F. A., Kumar, P., Shrivastava, G., & Bouhlel, M. S. (2018). Bitcoin: Digital Decentralized Cryptocurrency. In Handbook of Research on Network Forensics and Analysis Techniques (pp. 395-415). IGI Global.

Azzini, A., da Costa Pereira, C., Dragoni, M., & Tettamanzi, A. G. B. (2008, September). Evolving neural networks for word sense disambiguation. In *Hybrid Intelligent Systems, 2008. HIS'08. Eighth International Conference on* (pp. 332-337). IEEE. doi:10.1109/HIS.2008.88

Barber, S., Boyen, X., Shi, E., & Uzun, E. (2012, February). Bitter to better—how to make bitcoin a better currency. In *International Conference on Financial Cryptography and Data Security* (pp. 399-414). Springer. doi:10.1007/978-3-642-32946-3_29

Benincasa, C., Calden, A., Hanlon, E., Kindzerske, M., Law, K., Lam, E., & Whitaker, N. (2006). *Page Rank Algorithm. Department of Mathematics and Statics, University of Massachusetts*.

Cawrey, D. (2014). *Auroracoin Airdrop: Will Iceland Embrace a National Digital Currency?* CoinDesk.

Daly, E. M., & Haahr, M. (2007, September). Social network analysis for routing in disconnected delay-tolerant manets. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing* (pp. 32-40). ACM. doi:10.1145/1288107.1288113

Eisenmann, T. R., & Barley, L. (2006). *PayPal merchant services*. Academic Press.

Escudero, G., Màrquez, L., & Rigau, G. (2000). *Naive Bayes and exemplar-based approaches to word sense disambiguation revisited*. arXiv preprint cs/0007011

Farell, R. (2015). *An analysis of the cryptocurrency industry*. Academic Press.

Fei, B., & Liu, J. (2006). Binary tree of SVM: A new fast multiclass training and classification algorithm. *IEEE Transactions on Neural Networks*, *17*(3), 696–704. doi:10.1109/TNN.2006.872343 PMID:16722173

Grandjean, M. (2015). *GEPHI: Introduction to Network Analysis and Visualisation*. Academic Press.

Heard, N. A., Weston, D. J., Platanioti, K., & Hand, D. J. (2010). Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics*, *4*(2), 645–662. doi:10.1214/10-AOAS329

Huang, X., Shao, S., Wang, H., Buldyrev, S. V., Stanley, H. E., & Havlin, S. (2013). The robustness of interdependent clustered networks. *Europhysics Letters*, *101*(1), 18002. doi:10.1209/0295-5075/101/18002

Jack, W., & Suri, T. (2011). *Mobile money: The economics of M-PESA (No. w16721)*. National Bureau of Economic Research. doi:10.3386/w16721

Kalodner, H. A., Carlsten, M., Ellenbogen, P., Bonneau, J., & Narayanan, A. (2015, June). *An Empirical Study of Namecoin and Lessons for Decentralized Namespace Design*. WEIS.

Kennedy, A., & Szpakowicz, S. (2011, May). A supervised method of feature weighting for measuring semantic relatedness. In *Canadian Conference on Artificial Intelligence* (pp. 222-233). Springer. doi:10.1007/978-3-642-21043-3_27

King, S., & Nadal, S. (2012). Ppcoin: Peer-to-peer crypto-currency with proof-of-stake. Self-published paper.

Kuncheva, Z., & Montana, G. (2015). *Community detection in multiplex networks using locally adaptive random walks.* http://arxiv.org/abs/1507.01890

Leskovec, J., Krevl, A., & Datasets, S. N. A. P. (2014). *Stanford large network dataset collection, 2014.* http://snap. stanford. edu/data/index. html

Loera, A. (2015). *U.S. Patent Application No. 14/178,234*. US Patent Office.

Miller, A., Juels, A., Shi, E., Parno, B., & Katz, J. (2014, May). Permacoin: Repurposing bitcoin work for data preservation. In *2014 IEEE Symposium on Security and Privacy (SP)* (pp. 475-490). IEEE. doi:10.1109/SP.2014.37

Mittal, & Bhatia. (2018). Detection of Suspicious or Un-Trusted Users in Crypto-currency Financial Trading Applications. *International Journal of Digital Crime and Forensics*.

Mittal & Bhatia. (2019). Discovering Bottlenecks Entities in Multi-Layer Social Networks. JIOS/JSMS. (accepted)

Mittal, R., & Bhatia, M. P. S. (2017, July). Mining top-k structural holes in multiplex networks. In *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on* (pp. 1-6). IEEE. doi:10.1109/ICCCNT.2017.8204129

Mittal, R., & Bhatia, M. P. S. (2017). *Behaviour Analysis of Social Networks*. Accepted and Presented and published in International Conference on Computing for Sustainable Global Development (INDIACom).

Mittal, R., & Bhatia, M. P. S. (2018). Analysis of Multiplex Social Networks Using Nature-Inspired Algorithms. *Nature-Inspired Algorithms for Big Data Frameworks, 290*.

Mittal, R., & Bhatia, M. P. S. (2018). Anomaly Detection in Multiplex Networks. *Procedia Computer Science*, *125*, 609–616. doi:10.1016/j.procs.2017.12.078

Mittal, R., & Bhatia, M. P. S. (2018, July). Cross-Layer Closeness Centrality in Multiplex Social Networks. In *Computing, Communication and Networking Technologies (ICCCNT), 2018 9th International Conference on* (pp. 1-5). IEEE. doi:10.1109/ICCCNT.2018.8494042

Mittal, R., & Bhatia, M. P. S. (2018). Analyzing the Structures of Clusters in Multi-layer Biological Networks, *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 502-507. (Google Scholar) doi:10.1109/ICSCCC.2018.8703271

Mittal, R., & Bhatia, M. P. S. (2018). *Characterizing the Properties of the Multiplex Social Networks.* Presented and published in International Conference on Computing for Sustainable Global Development (INDIACom).

Mittal, R., & Bhatia, M. P. S. (2019) Identifying Prominent Authors from Scientific Collaboration Multiplex Social Networks. In *International Conference on Innovative Computing and Communications. Lecture Notes in Networks and Systems, vol 55*. Springer. doi:10.1007/978-981-13-2324-9_29

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. Academic Press.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256. doi:10.1137/S003614450342480

Newman, M. E. J. (2009). Random graphs with clustering. *Physical Review Letters*, *103*(5), 058701. doi:10.1103/PhysRevLett.103.058701 PMID:19792540

Niu, Z. Y., Ji, D. H., & Tan, C. L. (2005, June). Word sense disambiguation using label propagation based semi-supervised learning. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 395-402). Association for Computational Linguistics. doi:10.3115/1219840.1219889

Pinzón, C., & Rocha, C. (2016). Double-spend Attack Models with Time Advantange for Bitcoin. *Electronic Notes in Theoretical Computer Science*, *329*, 79–103. doi:10.1016/j.entcs.2016.12.006

Rogaway, P., & Shrimpton, T. (2004, February). Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In *International workshop on fast software encryption* (pp. 371-388). Springer.

Schwartz, D., Youngs, N., & Britto, A. (2014). *The Ripple protocol consensus algorithm*. Ripple Labs Inc White Paper.

Shao, S., Huang, X., Stanley, H. E., & Havlin, S. (2014). Robustness of a partially interdependent network formed of clustered networks. *Phys. Rev. E*, *89*(3), 032812. doi:10.1103/PhysRevE.89.032812 PMID:24730904

Shrivastava, G. (2017). Approaches of network forensic model for investigation. *International Journal of Forensic Engineering*, *3*(3), 195–215. doi:10.1504/IJFE.2017.082977

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

Tang, X., Wang, J., Zhong, J., & Pan, Y. (2014). Predicting essential proteins based on weighted degree centrality. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11*(2), 407-418.

Teutsch, J., Jain, S., & Saxena, P. (2016, February). When cryptocurrencies mine their own business. In *International Conference on Financial Cryptography and Data Security* (pp. 499-514). Springer.

Upadhyaya, R., & Jain, A. (2016, April). Cyber ethics and cybercrime: A deep dwelved study into legality, ransomware, underground web and bitcoin wallet. In *Computing, Communication and Automation (ICCCA), 2016 International Conference on* (pp. 143-148). IEEE.

Vasin, P. (2014). *Blackcoin's proof-of-stake protocol v2*. https://blackcoin. co/blackcoin-pos-protocolv2-whitepaper. pdf

Wadhwa, P., & Bhatia, M. P. S. (2012). *Social networks analysis: trends, techniques and future prospects*. Academic Press.

Wadhwa, P., & Bhatia, M. P. S. (2014). Community detection approaches in real world networks: A survey and classification. *International Journal of Virtual Communities and Social Networking*, *6*(1), 35–51. doi:10.4018/ijvcsn.2014010103

Wadhwa, P., & Bhatia, M. P. S. (2014). Classification of radical messages in Twitter using security associations. *Case studies in secure computing: Achievements and trends*, 273-294.

Wadhwa, P., & Bhatia, M. P. S. (2014). Discovering hidden networks in on-line social networks. *International Journal of Intelligent Systems and Applications*, *6*(5), 44–54. doi:10.5815/ijisa.2014.05.04

Wadhwa, P., & Bhatia, M. P. S. (2015). Measuring radicalization in online social networks using Markov Chains. *Journal of Applied Security Research*, *10*(1), 23–47. doi:10.1080/19361610.2015.972265

Wadhwa, P., & Bhatia, M. P. S. (2015). An approach for dynamic identification of online radicalization in social networks. *Cybernetics and Systems*, *46*(8), 641–665. doi:10.1080/01969722.2015.1058665

Wadhwa, P., & Bhatia, M. P. S. (2016). New Metrics for Dynamic Analysis of Online Radicalization. *Journal of Applied Security Research*, *11*(2), 166–184. doi:10.1080/19361610.2016.1137203

Watanabe, S., & Kabashima, Y. (2014). Cavity-based robustness analysis of interdependent networks: Influences of intranetwork and internetwork degree-degree correlations. *Phys. Rev. E*, *89*(1), 012808. doi:10.1103/PhysRevE.89.012808 PMID:24580282

Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper, 151*, 1-32.

Ruchi Mittal is a research scholar and is presently pursuing a Ph.D. in the Department of Computer Engineering at Netaji Subhas Institute of Technology, University of Delhi. Her areas of interest include Social network analysis, graph analysis, complex networks, multi-layer networks, digital forensics, and data mining.

M. P. S. Bhatia received his Ph.D. in Computer Engineering from the University of Delhi, Delhi. He has been working as a Professor in the Computer Engineering Department of Netaji Subhas Institute of Technology, New Delhi, India. He has guided many M.Tech and Ph.D. students. He has authored several papers in refereed journals, book chapters, and conferences. His research interests include Software Engineering, Cyber Security, and Privacy, Social Network Analysis, Data Mining, Machine Learning, Pattern Recognition, Semantic Web, Big Data and Analytics.