Ledgit: A Service to Diagnose Illicit Addresses on Blockchain using Multi-modal Unsupervised Learning

Xiaoying Zhi
xiazhi@ethz.ch
ETH Zurich, JPMorgan Chase
Switzerland

Yash Satsangi
yash.satsangi@jpmchase.com
JPMorgan Chase
United Kingdom

Sean Moran
sean.j.moran@jpmchase.com
JPMorgan Chase
United Kingdom

Shaltiel Eloul
shaltiel.eloul@jpmchase.com
JPMorgan Chase
United Kingdom

ABSTRACT

Distributed ledger technology benefits society by enabling an ecosystem of decentralised finance. However the pseudo-anonymised nature of transactions has also been an enabler of new routes for illicit activities ranging from individual scams to organised crimes. Current solutions for identifying addresses involved in illicit activities (illicit addresses) rely on commercial intelligence services, which are costly due to the intensive investigative efforts required. We propose *Ledgit*, an automatic real-time service for diagnosing illicit addresses on the Bitcoin blockchain. *Ledgit* is based solely on publicly available data, and uses an unsupervised clustering method that combines information from textual reports and the blockchain graph to assign a risk score that a Bitcoin address is involved in illicit activities. We verify the system with labeled addresses, showing high performance in identifying illicit addresses. Finally, we provide an intuitive user interface that provides accessible risk assessment with graph and report analytics.

CCS CONCEPTS

• **Applied computing** → *Secure online transactions*; • **Software and its engineering** → **Software design engineering**.

KEYWORDS

blockchain, risk, unsupervised learning, multimodality

ACM Reference Format:
Xiaoying Zhi, Yash Satsangi, Sean Moran, and Shaltiel Eloul. 2022. Ledgit: A Service to Diagnose Illicit Addresses on Blockchain using Multi-modal Unsupervised Learning. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557212

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557212

1 INTRODUCTION

Since the launch of the Bitcoin blockchain and its whitepaper in 2008 [24], distributed ledger technology (DLT) has grown rapidly. With this rapid growth there has been growing concern from individual and governmental bodies [2, 3, 12, 25] on how to effectively monitor pseudo-anonymous transactions to tackle organised crimes such as money laundering, ransomware, illicit cross-border fund movement [25] as instigated, for example, by terrorist and sanctioned groups and human/drug traffickers.

Subsequently much effort has been devoted to the monitoring of illicit activity (transactions that may support illegal activities) on crypto ledgers [18, 31]. With the large growth in the crypto-market, machine learning based models have became the center of various blockchain monitoring applications, such as of identity mixing [34], anomaly detection [14, 20, 21, 30], including various supervised learning models [17, 36], and some unsupervised models [20, 21, 29]. In 2019, Elliptic released the largest labeled dataset [33] on Bitcoin transaction activities, which contains a snapshot of a part of the Bitcoin blockchain transaction graph in time including their transaction features and mined labels (illicit/licit). This database triggered the development of a large number of supervised models that detect illicit transactions with high performance, utilising transaction information and their labels [9, 11, 16, 26, 28, 32, 33].

However it was recently shown that the dynamics of the blockchain plays a crucial role in the embedding of the graph-structured transactions, and using only historical data is insufficient to train a good real-time illicit classifier [16]. Therefore real-time and abundant mining of strong labels is necessary to train classifiers on blockchain ledgers. However mining such strong labels is costly, and hence there is hardly any automated real-time service that can leverage machine learning and assess the risk of n given Bitcoin address as illicit. In fact, the most practical and accessible way to measure the risk of an address for the end user is to rely on public reports from governments and recognised institutes. However those reports are updated very rarely and in a limited scope. An alternative is to check addresses in curated web services where public users can share votes and incidents [6, 7]. Public reports from platform such as "Bitcoinabuse.com" [6], are real-time and frequent, with over 200k reports with textual description of incidents and information such as categories, time *etc*. However it is a challenge to validate if these are genuine reports, mistakes, spam, fake, or advertisements.

Our solution and contributions. This paper proposes *Ledgit*, an automated real-time service that lets a user query the risk level of a Bitcoin address as licit or illicit. *Ledgit* uses an unsupervised learning algorithm that clusters addresses with similar reports and transactions together and computes the risk level of a Bitcoin address based on real-time publicly available data, namely the transaction graph data and the public reports from Bitcoinabuse.com. Ledgit uses textual features extracted from the Bitcoinabuse reports and 'behavioural' features from the reports' meta data. From the blockchain, Ledgit extracts graph features of the Bitcoin addresses in the transaction graph. Each Bitcoin address is then represented by the combination of the textual, behavioural and transaction graph features, where the textual features capture the topics and nature of complains, behavioural features capture the statistics such as number/frequency of reports, number of unique complainers and transaction graph captures information such as transaction frequency and volume as well as transactions with other addresses. Next, k-means [8, 19] is used to cluster the Bitcoin addresses and each address is assigned to categories of high, medium and low risk levels depending on resulting clusters and a carefully designed decision process. We validate our approach by using a list of known illicit addresses, labeled by governmental bodies. We use this algorithm and clusters in a back-end to support a user friendly front-end interface where a user can input a Bitcoin address to check its risk level and analyse its graph and associated reports. In the rest of the paper we first demonstrate the user interface service followed by our clustering algorithm and main results.

2 OVERVIEW OF DEMO SYSTEM

Figure 1a shows the general workflow of the service. The user's input is a Bitcoin/wallet address. The user interface (UI) includes a pop-up query to input a new address. Upon receiving an address, the service samples a mini-graph by crawling the blockchain around the queried address, then returns a Bitcoin flow directed mini-graph that captures the transactions and interactions of the given address and its neighbors, as shown in Figure 1b. This output UI consists of the following parts: the sampled mini-graph, with the queried address as the center, an analytics table that displays the information extracted from Bitcoinabuse reports, and the individual risk predictions. The output mini-graph shows addresses as nodes and the flow (transactions) of Bitcoin as directed edges. Dominant addresses and transactions (large flow of Bitcoins) are proportionally magnified to expose dominant paths. Additionally, if a certain address (node) in the graph is listed in any authority/trusted/community lists, there is an indicator around it mentioning the source of listing. The risk predictions from Bitcoinabuse reports are displayed as a standalone analytics table over the canvas. The analytics table also displays the number of reports associated with an address, the time of the latest report, the breakup of risks based only on Bitcoinabuse reports and transaction information. Finally, the displayed mini-graph is an interactive one, meaning the user can drag, zoom, rotate, for viewing an information box of each node when the pointer floats over it. The backend process which carries out the computation and the risk decision process is discussed next in Section 3.

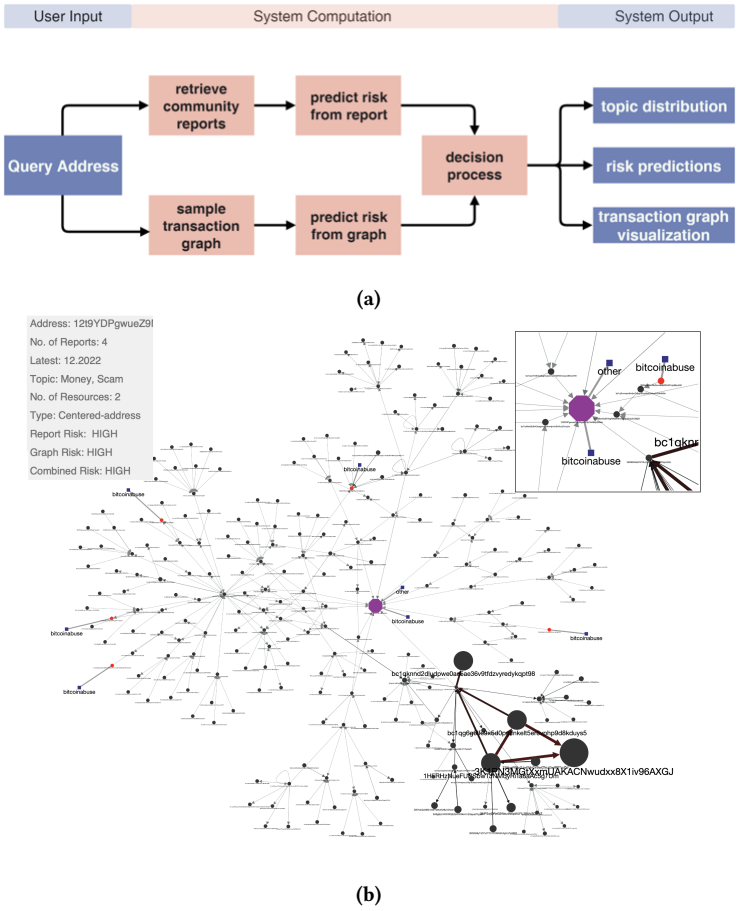


Figure 1: (a) Overall workflow of the web service (b) Snapshot from the user interface for a query of an address. The output mini-graph shows addresses as nodes and the flow (transactions) of Bitcoin as edges. Prediction of the risk and analytics are provided for the central address. Addresses with illicit labels or with reports from Bitcoinabuse are highlighted in red with attached information on each resource (blue squares). Dominant paths are magnified. The inset shows a close-up on the centered address.

3 SYSTEM ARCHITECTURE AND CLUSTERING ALGORITHM

Figure 2 shows how *Ledgit* generates the three risk diagnoses for an address. The system firstly uses k-means to cluster the constructed features from the two data sources (report and graph) individually, and then we assign the associated risk level to each cluster. To predict the risk for a new address, Ledgit uses k-nearest neighbors (KNN) [13] to map the new address to existing clusters and then follows a decision process to obtain the combined risk.

3.1 Dataset and Feature Building

Ledgit uses two data sources, public reports and transaction graphs.

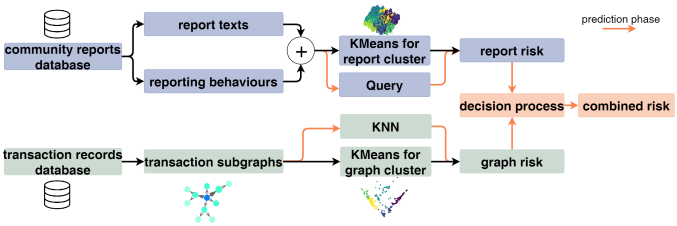


Figure 2: An illustration of the system architecture. In the clustering phase each data type goes through feature building, k-means, and risk assignment. In prediction, each data type goes through the same feature building, KNN on query, risk lookup, and a combined decision process.

Public reports. We use the publicly available report dataset from Bitcoinabuse.com containing all reports until the day of retrieval. Each report consists of a textual description of the suspicious activity reported by a victim. Additionally there is abundant metadata such as the date and time, the country, abuse type, user name *etc.* However a significant portion of these reports are simply spam, advertisement or addresses that are redirected to exchange nodes. We tackle the challenge of filtering spam in these reports with the help of insights from the existing literature on unsupervised opinion spam detection [22, 23]. These methods work by identifying key behavioural features (example: number of reviews) and textual information (content similarity) that seem to help with unsupervised spam detection [22, 23]. Based on these insights we extract behavioural features from the metadata such as the latest report time, average report time, the number of reports and number of unique reporters on each address. In order to extract textual features we make use of the universal sentence encoder (USE) [35] to encode the textual reports into vector representations. Finally, we also add features that represent the topic distribution from the collected report texts among some pre-defined topics (such as malware, website, sextortion, etc). We extracted these topics from each report using the non-negative matrix factorisation (NMF) model [15]. Overall, the text embedding, the topics one-hot vectors and the behavioural features are concatenated as the feature embedding of the report data, resulting in a feature vector of length 525, which followed by a k-means clustering.

Mini-Graphs. Mini-graphs of the addresses transactions are sampled by crawling on the Bitcoin blockchain [1] itself around a centered address of interest in real-time by using the blockchain API. We use a radius of up to 3 hops from the address. In the graphs each node represents an address and each edge represents a directed flow of Bitcoin from one address to another. We limit the crawler by choosing the top 5 highest rates of transaction for each neighbour node. We construct a set of features: a) *structure-related features e.g.* number of edges and nodes, number of addresses at each distance from center, b) *transaction-related features e.g.* value of each transaction, node balance, and c) *disclosure-related features e.g.* number of nodes at each distance which are disclosed by any report or labeled list. The feature values are normalised and form a 34-dimensional feature vector.

Verification dataset. For verification we construct a labeled dataset with both licit and illicit addresses for verification of the report and graph clustering. The illicit part is a combination of government disclosure [4, 5] and a publicized ransomware dataset [27]. The licit addresses can be extracted from the de-anonymised licit transactions in the Elliptic dataset [33], based on the rationale that a licit transaction should only involve licit addresses. Note that it is impossible to assume the same for illicit transactions, hence for illicit labels, we rely on the limited publicly trusted resources mentioned above. For the verification of graph clustering, we sampled 197 illicit and 156 licit transactions. For verification of the reports clustering and the graph clustering together (Section 4.1 and 4.2), we use 51 illicit and 24 licit addresses that were found to have reports.

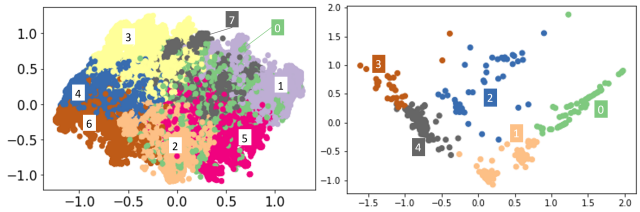
3.2 Clustering and Prediction

For report clustering, we use the entire unlabeled report dataset with 81, 629 addresses and 266, 028 reports. For graph clustering, we use the sampled 197 illicit and 156 licit addresses with masked labels, covering 353 addresses and 132, 564 transaction records. Once we have the feature vectors from public reports and transaction graphs, we use k-means clustering to cluster each set of features independently to prevent the features from one modality dominating the clustering. We explored various values of k in between 0 and 24 and chose the number of clusters as k=8 and k=5 for report and graph clustering since it lead to clearer separation of clusters, for example, address with longer life span were separated from once with shorter life span. Further experimentation with other algorithms [10] are possible but planned as future work since focus of this paper is to build a real-life system.

4 CLUSTERING RESULTS

Figure 3a shows the distribution of the 8 report clusters and Table 1 provides several average features associated with the clusters. Initially, the risk levels are assigned by a simple human inspection of average key features. Then, following verification dataset, the risk levels for the clusters are further refined to maximise recall of illicit addresses. A key result of clustering as highlighted by Table 1 is a clear separation between address with high number of reports (cluster 5: likely to contain genuine reports) and addresses that have nominal reports against them (likely to be spam). Specifically, cluster 5 corresponds to addresses with high number of reports and longer lifetime and is well separated from other clusters with average number of around 3 reports or less. Notably, clusters with similar number of reports were in some cases separated depending on their duration on the website (old or new) or textual information of reports in those clusters. Hence, we assign cluster 5 and 3 (top 2 number of reports) with a HIGH risk level, and cluster 0 and 4 a risk of LOW and the rest MEDIUM.

Note that a high number of unique reports does not always relate to illicit reports. For example, many reports may direct to an address that belongs to an exchange platform. This is an example why the combination with graph information is useful, since exchange nodes have a unique topology with high graph degree. Figure 3b visualises the distribution of the 5 graph clusters. For graph clustering (k=5), we assign risk level by observing anomaly of large



(a) (b)

Figure 3: Feature distribution of each cluster in space, dimensioned reduced by PCA. (a) report clusters; (b) graph clusters.

Table 1: Cluster profiles of reports

cluster	avg. no. of reports	avg. duration (hour)	% of addresses
0	1.42	140.78	5.80
1	**3.02**	**210.42**	11.78
2	1.96	98.92	18.24
3	1.83	**218.51**	9.36
4	1.95	118.97	15.95
5	**9.05**	**544.57**	14.85
6	2.42	87.50	9.31
7	2.32	157.72	14.71

Table 2: Cluster profiles of graphs. Transaction (txn) values are measured in Bitcoin.

cluster	avg. txn value	avg. txn std	% of addresses
0	**104.65**	**880.90**	20.11
1	6.51	37.48	24.93
2	**37.57**	**386.94**	12.18
3	11.58	77.51	8.22
4	3.30	16.80	34.56

average transaction values (Table 2), and label the two clusters with excessively high average transaction values, either total, incoming, or outgoing, as HIGH risk (clusters 0 and 2). The lowest average value is assigned as LOW (cluster 4).

4.1 Clustering Verification and Decision Process

For verification we sampled a subset of addresses from the verification dataset (as discussed in Section 3.1), including 51 illicit and 24 licit addresses. The results of classification are shown in Table 3. After evaluation of the addresses, we refined the initial cluster risk labels for report of cluster 2 as it contains 24% of all illicit samples. Results show that our assignments of risk levels from graphs and reports overall provides a reasonable assessment of risk that can be used for monitoring new unknown addresses. We further provide a combined risk using a decision process based on our analysis.

4.2 Combined Decision Process

Figure 4 illustrates the combined decision process for risk assignment. We simplify the final decision to three categories: HIGH, LOW, and UNKNOWN. The decision prioritises graph clustering since it outperforms the reports clustering in detecting illicit addresses. In the case that the assigned risk by graph clustering is MEDIUM/LOW, addresses that have reports are assigned firstly by their report risk. Otherwise, if they do not have reports we classify their risk by the presence (as a ratio) of trusted disclosures (*e.g.* [4, 5, 27]) in the address mini-graph. Finally, the combined risk shows high F1-score of 93%, a high performance even when qualitatively compared to supervised non-real-time studies [33]. We thus provide this combined risk in our service in addition to the individual risks from the report and graph clustering.

Table 3: Evaluation of three clustering methods. All metrics are calculated under the setup that 'illicit' or 'HIGH risk level' corresponds to 'Positive' class.

	report risk	graph risk	combined risk
accuracy	0.7467	0.8400	**0.9067**
precision	0.9444	**0.9756**	0.9400
recall	0.6667	0.7843	**0.9215**
F1-score	0.7816	0.8695	**0.9306**

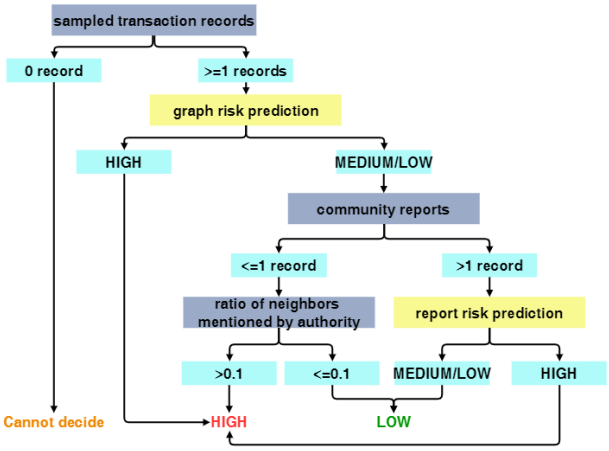


Figure 4: The decision process for combined risk.

5 CONCLUSION

We present *Ledgit*, an automatic system that can diagnose risk on an unknown Bitcoin address by only using public information. The system is the first to mine and combine information from public reports and blockchain transaction graphs to provide the risk of illicit activity by Bitcoin addresses. This system extended in the future to more resources and crypto-assets.

ACKNOWLEDGEMENTS

We thank Alex Stoliar, Helene Khaykovich, and Shravan Parunandula for the fruitful discussions and their feedback. We thank the anonymous reviewers for their feedback.

REFERENCES

[1] 2008. Blockchain.com. https://www.Blockchain.com/

[2] 2019. Financial Conduct Authority, Joint Money Laundering Steering Group (JMLSG) PART II, Chapter 22. https://www.fca.org.uk/firms/financial-crime/cryptoassets-aml-ctf-regime

[3] 2020. European Commission, Directorate-General for Financial Stability, Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Markets in Crypto-assets, and amending Directive (EU). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020PC0593

[4] 2022. . National Bureau for Counter Terror Financing of Israel. https://nbctf.mod.gov.il/en/seizures/Pages/Blockchain1.aspx

[5] 2022. . United States Office of Foreign Assets Control. https://www.treasury.gov/ofac/downloads/sdnlist.pdf

[6] 2022. BitcoinAbuse. https://www.bitcoinabuse.com/

[7] 2022. Scam-alert.io. https://scam-alert.io

[8] Charu C Aggarwal and Chandan K Reddy. 2014. Data clustering. *Algorithms and applications. Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra* (2014).

[9] Ismail Alarab, Simant Prakoonwit, and Mohamed Ikbal Nacer. 2020. Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain. In *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*. 23–27.

[10] Humam Alwassel, Dhruv Mahajan, Bruno Korbar, Lorenzo Torresani, Bernard Ghanem, and Du Tran. 2020. Self-supervised learning by cross-modal audio-video clustering. *Advances in Neural Information Processing Systems* 33 (2020), 9758–9770.

[11] Massimo Bartoletti, Barbara Pes, and Sergio Serusi. 2018. Data Mining for Detecting Bitcoin Ponzi Schemes. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. 75–84. https://doi.org/10.1109/CVCBT.2018.00014

[12] Joseph R. Biden. 2022. Executive order on ensuring responsible development of Digital assets. https://www.whitehouse.gov/briefing-room/presidential-actions/2022/03/09/executive-order-on-ensuring-responsible-development-of-digital-assets/

[13] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

[14] Ramiro Daniel Camino, Radu State, Leandro Montero, and Petko Valtchev. 2017. Finding Suspicious Activities in Financial Transactions and Distributed Ledgers. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. 787–796. https://doi.org/10.1109/ICDMW.2017.109

[15] Andrzej CICHOCKI and Anh-Huy PHAN. 2009. Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* E92.A, 3 (2009), 708–721. https://doi.org/10.1587/transfun.E92.A.708

[16] Shaltiel Eloul, Sean J Moran, and Jacob Mendel. 2021. Improving Streaming Cryptocurrency Transaction Classification via Biased Sampling and Graph Feedback. In *Annual Computer Security Applications Conference* (Virtual Event, USA) *(ACSAC)*. Association for Computing Machinery, New York, NY, USA, 761–772. https://doi.org/10.1145/3485832.3485913

[17] Yining Hu, Suranga Seneviratne, Kanchana Thilakarathna, Kensuke Fukuda, and Aruna Seneviratne. 2019. Characterizing and Detecting Money Laundering Activities on the Bitcoin Network. arXiv:1912.12060 [cs.SI]

[18] Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse. 2021. Knowledge Discovery in Cryptocurrency Transactions: A Survey. *IEEE Access* 9 (2021), 37229–37254. https://doi.org/10.1109/ACCESS.2021.3062652

[19] Stuart Lloyd. 1982. Least squares quantization in PCM. *IEEE transactions on information theory* 28, 2 (1982), 129–137.

[20] Patrick Monamo, Vukosi Marivate, and Bheki Twala. 2016. Unsupervised learning for robust Bitcoin fraud detection. In *2016 Information Security for South Africa (ISSA)*. IEEE, 129–134.

[21] Patrick M. Monamo, Vukosi Marivate, and Bhesipho Twala. 2016. A Multifaceted Approach to Bitcoin Fraud Detection: Global and Local Outliers. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 188–194. https://doi.org/10.1109/ICMLA.2016.0039

[22] Arjun Mukherjee. 2015. Detecting deceptive opinion spam using linguistics, behavioral and statistical modeling. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing: Tutorial abstracts*. 21–22.

[23] Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 632–640.

[24] Satoshi Nakamoto. 2008. *Bitcoin: A peer-to-peer electronic cash system*. Technical Report. Manubot.

[25] Emily Nicolle. 2022. U.K. Crime Agency Wants to Regulate Crypto Transaction Mixers. https://www.bloomberg.com/news/articles/2022-03-15/u-k-crime-agency-wants-to-regulate-crypto-transaction-mixers

[26] Michał Ostapowicz and Kamil Zbikowski. 2019. Detecting Fraudulent Accounts on Blockchain: A Supervised Approach. In *Web Information Systems Engineering – WISE 2019*, Reynold Cheng, Nikos Mamoulis, Yizhou Sun, and Xin Huang (Eds.). Springer International Publishing, Cham, 18–31.

[27] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoit Dupont. 2018. Ransomware Payments in the Bitcoin Ecosystem. https://doi.org/10.5281/zenodo.1238041

[28] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. 2019. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. arXiv:1902.10191 [cs.LG]

[29] Thai Pham and Steven Lee. 2016. Anomaly detection in bitcoin network using unsupervised learning methods. *arXiv preprint arXiv:1611.03941* (2016).

[30] Thai Pham and Steven Lee. 2017. Anomaly Detection in the Bitcoin System - A Network Perspective. arXiv:1611.03942 [cs.SI]

[31] Dorit Ron and Adi Shamir. 2013. Quantitative analysis of the full bitcoin transaction graph. In *International Conference on Financial Cryptography and Data Security*. Springer, 6–24.

[32] Dylan Vassallo, Vincent Vella, and Joshua Ellul. 2021. Application of Gradient Boosting Algorithms for Anti-money Laundering in Cryptocurrencies. *SN Computer Science* 2, 3 (2021), 1–15.

[33] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I. Weidele, Claudio Bellei, Tom Robinson, and Charles E. Leiserson. 2019. Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. https://doi.org/10.48550/ARXIV.1908.02591

[34] Jiajing Wu, Jieli Liu, Weili Chen, Huawei Huang, Zibin Zheng, and Yan Zhang. 2021. Detecting mixing services via mining bitcoin transaction network with hybrid motifs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2021).

[35] Yinfei Yang, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2019. Multilingual Universal Sentence Encoder for Semantic Retrieval. https://doi.org/10.48550/ARXIV.1907.04307

[36] Haohua Sun Yin and Ravi Vatrapu. 2017. A first estimation of the proportion of cybercriminal entities in the bitcoin ecosystem using supervised machine learning. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 3690–3699.