frontiers in Computer Science

REVIEW
published: 30 November 2020
doi: 10.3389/fcomp.2020.600596

Analysis Techniques for Illicit Bitcoin Transactions

Adam Brian Turner, Stephen McCombie and Allon J. Uhlmann*

Department of Security Studies and Criminology, Macquarie University, Sydney, NSW, Australia

This comprehensive overview of analysis techniques for illicit Bitcoin transactions addresses both technical, machine learning approaches as well as a non-technical, legal, and governance considerations. We focus on the field of ransomware countermeasures to illustrate our points.

Keywords: bitcoin, machine learning, graph analysis, financial crime, ransomware, AML

INTRODUCTION

This paper examines the current literature on the analysis of illicit Bitcoin transactions and focuses specifically on the analytic techniques that are applied to blockchain data. These illicit Bitcoin transactions could take the form of money laundering, terrorism financing or the movement of proceeds from other crimes such as ransomware attacks. Many of the techniques wrestle with the problem of attribution in the face of the anonymity of sources within the Bitcoin ecosystem. Therefore, we first examine the body of literature relating to regulatory efforts that aim to balance the freedom of an open system with the requirements of crime prevention and law enforcement. Following that is a review of the research into the techniques that exploit heuristics and behaviors inherent in the Bitcoin system. We then highlight the application of graph analysis techniques to the Bitcoin ecosystem and transaction networks. Furthermore, Machine Learning (ML) and Artificial Intelligence (AI) techniques applied to money laundering, cybercrime and other illicit activities across the Bitcoin ecosystem are reviewed. Moreover, a focus is placed on the application of these techniques to the modern day threat of ransomware, a lucrative branch of contemporary global crime which in 2020 is estimated to cost companies anywhere between $US 42 billion and $US 170 billion worldwide in ransoms paid, lost productivity and other recovery expenses (Emsisoft, 2020).

REGULATORY AND COMPLIANCE CHALLENGES

The regulatory landscape has continuously evolved since Nakamoto (2008) released the inaugural paper on Bitcoin, "*A peer-to-peer electronic cash system*" (Nakamoto, 2008). The decentralized nature of the peer to peer network from which Nakamoto (2008) designed Bitcoin affords the user anonymity and bypasses the central authority used to regulate traditional financial systems.

The Regulatory Environment

Tsukerman (2015) surveys the state of the Bitcoin regulatory environment from a United States (US) centric position. To help understand this environment they provide a breakdown of the laws into two categories: those laws that protect consumers who use Bitcoin; and those that address the broader societal impacts of people using Bitcoin for illegal purposes such as money laundering and terrorist financing (Tsukerman, 2015). Tu and Meredith (2015) complement the work by Tsukerman (2015) by considering the impediments to effective regulation of Bitcoin which addresses the issues of ownership, attribution and the susceptibility to theft, that virtual currencies

OPEN ACCESS

Edited by:
Alex C. Ng,
La Trobe University, Australia

Reviewed by:
José Antonio Álvarez-Bermejo,
University of Almeria, Spain
Gabriele Costa,
IMT School for Advanced Studies
Lucca, Italy

****Correspondence:***
Adam Brian Turner
adam.turner@students.mq.edu.au

Specialty section:
This article was submitted to Computer Security, a section of the journal Frontiers in Computer Science

Received: *30 August 2020*
Accepted: *10 November 2020*
Published: *30 November 2020*

Citation:
Turner AB, McCombie S and Uhlmann AJ (2020) Analysis Techniques for Illicit Bitcoin Transactions. Front. Comput. Sci. 2:600596. doi: 10.3389/fcomp.2020.600596

are subject to. Wagstaff and Karpeles (2014) reported on the largest theft of Bitcoin at the Bitcoin exchange Mt Gox in February 2014. This breach saw the exchange lose 850,000 Bitcoins worth $US 450 million at the time. Reclamation of these stolen funds is identified as a major risk to users by Tu and Meredith (2015). Irwin and Turner (2018) argue that cryptocurrency systems, in contrast with traditional money transmission businesses and financial institutions, are relatively unhindered by anti-money laundering and counter-terrorism financing (AML/CTF) regulations. In addition, these systems do not collect the necessary Personal Identifiable Information (PII) that will allow for the implementation of strict financial transaction reporting procedures for the purposes of mitigating illicit financial activity and the misappropriation of funds (Irwin and Turner, 2018). The procedures discussed in Irwin and Turner (2018) aim to examine the atypical business dealings conducted over Bitcoin, along with the use of AML/CTF techniques potentially indicating illicit activity.

In June 2018, The Law Library of Congress (2018) published a paper on "*Regulation of Cryptocurrency in Selected Jurisdictions.*" This report provides a comprehensive review of the cryptocurrency regulation and policy stance of the following jurisdictions: Argentina, Australia, Belarus, Brazil, Canada, China, France, Gibraltar, Iran, Israel, Japan, Jersey, Mexico, Switzerland. For each of these jurisdictions, there is a foreign law specialist assigned to assess the legal conditions within the respective jurisdiction. During the introduction of this report, foreign law specialist Hanibal Goitom identifies the major issues jurisdictions are facing. Namely, the legality of cryptocurrency operations, issues around taxation and AML/CTF implications.

Legality of Cryptocurrency Markets

By revealing how different countries are legally operating cryptocurrency markets in their jurisdictions the report highlights specific laws enacted for cryptocurrency markets to operate and the contrasting jurisdictions that restrict their trade. It identifies the likes of Belarus, Gibraltar, Jersey, and Mexico have enacted laws specifically recognizing cryptocurrency markets. For example, in Belarus The Presidential Decree on the Development of the Digital Economy initiated on March 28, 2018 provides a legal framework for "*buying, selling, exchanging, creating, and mining cryptocurrencies and tokens.*" (Decree of the President of the Republic of Belarus No. 8., 2017). The Decree sets out a specific economic zone for companies to operate cryptocurrency related exchanges and services. In contrast countries such as China and Iran are excluding financial institutions within their jurisdiction from engaging in cryptocurrency markets. For instance, Pilarowski and Yue (2017) identify eight entities in China providing governance and oversight on the prevention of cryptocurrency usage. These entities are: "*the People's Bank of China (PBOC), the Cyberspace Administration of China (CAC), the Ministry of Industry and Information Technology (MIIT), the State Administration for Industry and Commerce (SAIC), the China Banking Regulatory Commission (CBRC), the China Securities Regulatory Commission (CSRC), and the China Insurance Regulatory Commission (CIRC).*" (Pilarowski and Yue, 2017). They all announced a ban on Initial Coin Offerings (ICOs) on September 4, 2017. The reason sighted was down to investor protection and financial risk prevention (Pilarowski and Yue, 2017).

Taxation

Tax evasion is an important but peripheral topic to this paper, however, Goitom, from The Law Library of Congress (2018) highlights the issue of how cryptocurrencies are taxed across various jurisdictions. This is a wide-ranging debate on the application of Tax Law against how cryptocurrencies are treated as a financial instrument. The Tax debate falls outside the scope of this review.

Anti-money Laundering (AML)/Counter Terrorism Financing (CTF)

The spring 2020 Cryptocurrency Crime and Anti-Money Laundering report from blockchain intelligence and forensics company CipherTrace revealed the global amount of Bitcoin crime attributed to fraud and misappropriation as $US 4.5 billion in 2019 (CipherTrace, 2020). A high proportion of these illicit Bitcoin transactions (74%) moved from exchange-to-exchange across jurisdictional borders. The report argues that the nature of these "cross-border" transactions emphasizes the need for cryptocurrency exchanges to adopt and ensure appropriate AML and CTF compliance is achieved. Efforts to regulate this in the Bitcoin context are evident in the AML laws, regulation and compliance instruments such as The Anti-money Laundering (AML) and Counter-terrorism Financing (CTF) Act 2006 (Cth) in Australia (Anti-Money Laundering and Counter-Terrorism Financing Act 2006. No. 169, 2006). The Australian AML/CTF Act calls for reporting entities to verify a customer's identity before the provision of a designated service (see Section 6 of the AML/CTF Act). In addition, risks need to be individually assessed for specific types of services and customers, how these services will be delivered to the customers, any foreign jurisdictions being traversed, and the state of connection of any financial entity performing a service in a foreign jurisdiction. In addition, the 5th Anti-Money Laundering Directive of the European Union (EU, 2018) provides a legislative framework for the prevention and detection of money laundering and terrorism financing in virtual currencies and exchanges. The EU directive (2018) places an emphasis on the national Financial Intelligence Units (FIUs) to "*combat the risks related to the anonymity,*" and that the FIUs "*should be able to obtain information allowing them to associate virtual currency addresses to the identity of the owner of virtual currency.*" (EU, 2018. Section 9). Provisions under these AML/CTF regimes define standards on Know Your Customer (KYC) and Customer Due Diligence (CDD) processes. Financial institutions and FIUs can leverage stringent KYC and CDD practices to enable essential customer identification procedures for a reporting entity. Irwin and Turner (2018) emphasize KYC and CDD as critical for linking the real-world identity of a customer's behavior and developing an understanding of their expected financial activities. Furthermore, to counter any AML/CTF risks, KYC and CDD ultimately satisfy the legal obligations to protect consumers and society from any misuse of virtual currencies for criminal purposes.

Financial Intelligence Units

Supporting these legislative frameworks are prominent FIUs such as the Financial Crimes Enforcement Network (FinCEN), The Financial Action Task Force (FATF) and the Australian Transaction Reports and Analysis Centre (AUSTRAC). FinCEN is the FIU of the US Treasury supporting US and international law enforcement investigations. In addition, FinCEN issues guidance and advisory notices regarding illicit usage of virtual currencies (FinCEN, 2019). For example, FIN-2019-G001 (2019), *Application of FinCEN's Regulations to Certain Business Models Involving Convertible Virtual Currencies*, is a comprehensive guidance to persons engaging with money services businesses (MSBs) that involve the transmission of convertible virtual currencies (CVCs) and how they are subject to the US Bank Secrecy Act. FIN-2019-G001 (2019) provides the necessary definitions and applications of the Bank Secrecy Act along with the obligations required when dealing with CVCs.

FATF provides recommendations and standards for over 200 jurisdictions to help prevent money laundering and terrorism financing. The FATF secretariat is located at the OECD Headquarters in Paris. The FATF International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation–*the FATF Recommendations* provide a comprehensive and consistent framework of measures allowing countries to implement to fight against money laundering and terrorist financing (FATF, 2012). Within that framework there are provisions explicitly relating to Virtual Assets (VAs) and Virtual Asset Service Providers (VASPs). The guidance document for a *Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers* (FATF, 2019) identifies "*risk indicators that should specifically be considered in a VA context, with an emphasis on factors that would further obfuscate transactions or inhibit VASPs' ability to identify customers.*" (FATF, 2019). Furthermore, it enhances the original FATF recommendations (FATF, 2012) amending FATF Recommendation 15 by requiring "*VASPs be regulated for anti-money laundering and combating the financing of terrorism (AML/CFT) purposes, licensed or registered, and subject to effective systems for monitoring or supervision.*" (FATF, 2019).

AUSTRAC is Australia's primary financial intelligence agency and has primary responsibility for AML/CTF intelligence collection and analysis. In addition, it provides guidance to entities against the Anti-Money Laundering and Counter-Terrorism Financing Act 2006 and the Financial Transaction Reports Act 1988. AUSTRAC manages the register for digital currency exchange businesses in Australia, along with a guide to preparing and implementing an AML/CTF program for digital currency exchange businesses (AUSTRAC, 2019).

Clearly, law enforcement agencies lack globally consistent procedures, laws, regulations or standards to police the misuse of cryptocurrencies. The FATF strives to set out global standards to combat money laundering and terrorist financing, and other significant threats that exist to disrupt the integrity of the global financial system. However, in most countries, when it comes to cryptocurrency operators there is no enforcement of the "know your customer" procedures or the intention to validate the identity of customers undertaking cryptocurrency transactions. According to The Law Library of Congress (2018) a number of countries are beginning to look at regulating cryptocurrencies and formulating policy frameworks. Furthermore, CipherTrace (2020), highlight the potential effectiveness of AML measures by indicating a 47% drop in criminal funds being sent directly to exchanges. Albeit a subjective link, CipherTrace suggest that this could be down to the AML controls inhibiting the exchange or cash-out of illicit proceeds.

This along with the EU directive (2018), underlines the significance of enabling authorities to monitor the use of virtual currencies. By authorizing FIUs to monitor the use of cryptocurrencies, the EU directive (2018) provides a step toward a more holistic approach for entities to combat the AML/CFT threat. The directive further states, "*Such monitoring would provide a balanced and proportional approach, safeguarding technical advances and the high degree of transparency attained in the field of alternative finance and social entrepreneurship.*" (EU, 2018. Section 8).

Challenges remain anchored in the international nature of cryptocurrency transactions and any resultant cybercriminal activity. To counter this challenge, it will be essential to prevent offenders from hopping from one jurisdiction to another. To impede such behaviors the enforcement of AML/CTF KYC provisions will act as a deterrent. The application of more stringent provisions could risk stifling the innovative functionality of cryptocurrencies, but at the same time balance out any illicit usage by having the capability to reveal the true identity of those participating in cryptocurrency. However, for the tradeoffs to be effective international cooperation, information sharing and monitoring between law enforcement agencies, FIUs and cryptocurrency service providers will be required.

This type of monitoring demands analysis techniques based on graph theory and network analysis which can produce predictive features and a machine learning architecture to manage large datasets. Implementation of machine learning architectures is intended to improve monitoring and investigations over time and would be less manpower intensive. In the next section we will review the literature pertaining to such techniques.

BITCOIN ANALYSIS

In the Beginning

After the release of the Nakamoto (2008) whitepaper, *A peer-to-peer electronic cash system. Bitcoin*, the early analysis of Bitcoin revolved around understanding the mechanics of the system. This is evident in Kaminsky (2011) who presented findings on the interaction of the Bitcoin protocol with Internet security protocols. In addition, Rosenfeld (2011) examined how the mining process works in order to reward participants on the Bitcoin network, Karame et al. (2012) looked at the "double spending attack" examining how to take advantage of the early stage Bitcoin transaction processing times and Drainville (2012) looked at the privacy motivations for using Bitcoin along with attack vectors that aim to compromise security and anonymity of the Bitcoin system. Then Stokes

(2012), broke ground on the utility of virtual currencies applied to money laundering. However, Reid and Harrigan (2011), Ron and Shamir (2012), and Meiklejohn et al. (2013), pioneered the fundamental techniques for analyzing Bitcoin transaction behavior.

Bitcoin Heuristics

Investigation into illicit Bitcoin usage creates a mosaic of information that must be forensically reconstructed to provide an accurate view of the target. The information can be technological, behavioral, criminological and regulatory in nature. The introduction of heuristics into the analysis can help address the difficulties of attribution. This is achieved by grouping similar transactional behavior and linking ownership to addresses and services on the Bitcoin network.

Meiklejohn et al. (2013) produced a seminal paper on analyzing the Bitcoin blockchain to reveal identity. The heuristics presented within this paper form the basis of which much of today's Bitcoin analysis is performed. This work makes it possible to cluster activity around a certain user and add context to this user for purposes of identification or grouping similar services on the network. In addition, it introduces the concept of peeling, where smaller amounts of Bitcoin are "peeled" off a larger amount and transferred onto another address with the remainder transferred back to the one-off change address. In addition, they discover, if a user of an input address also controls a one-off change address associated to that transaction, it may be assumed that both addresses are owned by the same user. This common pattern can be used to obfuscate the movement of funds and result in the detection of money laundering on the Bitcoin network. Meiklejohn et al. (2013) produce various other time-series analyses along with Bitcoin service breakdown analysis to understand and model the effects of the different services on the Bitcoin network. Meiklejohn et al. (2013), apply this type of analysis on aggregated data to help profile and characterize different activity trends on the Bitcoin network. Drilling deeper into the payment trends allows for a more targeted understanding of illicit user activity, especially its source. They also determined that it was only possible to identify ownership after any suspicious activity had occurred. Predicting that suspicious activity is going to take place in the future requires the collection of targeted Bitcoin addresses or transaction IDs to learn and train models for future prediction, investigation and analysis. Therefore, there is a need to look at other information sources to determine possible fraudulent transactions. This is where Reid and Harrigan (2011) posited cluster analysis as a technique to reveal patterns, associations, structures and relationships emanating from different data sources. Clustering can be used to identify common entities on the Bitcoin network controlling Bitcoin addresses by building up a picture of transaction flows over time. Nakamoto (2008), implies that clustering algorithms can group together multiple input transactions controlled by the same address, potentially identifying the owner of the address (Nakamoto, 2008). This makes it possible to construct a user network identifying mappings between Bitcoin addresses and a cluster of similar users (Reid and Harrigan, 2011). There is also the potential to find connection between Bitcoin addresses, IP addresses and spending patterns through this type of analysis.

Analyzing the Network Layer

To de-anonymize users on the Bitcoin network, Turner and Irwin (2018) look at the openness of the Bitcoin system and some of the defining features seen within the anatomy of a Bitcoin transaction coupled with extensive data collection from packet sniffing software. Using network traffic analyzer tools, such as Wireshark, can capture Bitcoin protocol traffic by listening on the network to port 8333 and building a profile of transaction flow between IP addresses and Bitcoin addresses over time. This is known as public key profiling. This method has weaknesses, such as the potential of Bitcoin addresses to change as frequently as every transaction. If this is the case, it will result in weak linkages to any network observations. Due to the peer-to-peer propagation of transactions any observation of an IP address where a transaction is intercepted may not be the original creator of the transaction. This further removes any ability to reveal identity *via* Bitcoin address usage analysis on the network (Turner and Irwin, 2018). Furthermore, Irwin and Turner (2018), highlight the lack of reliability in this analysis approach and the inhibitors of revealing any illicit transaction. They state: "*IP addresses that connect to computers in a library, café, open wireless network, virtual private network or Tor exit relay, used by many people, do not identify the perpetrator and, therefore, is not probable cause that a person was responsible for the communication or illicit activity.*" (Turner and Irwin, 2018). Nakamoto (2008) designed the Bitcoin system so that actors are pseudonymous. In addition, the transaction packet moving through the Bitcoin network does not contain the IP address. Only transaction IDs are ultimately stored on the blockchain. The transaction payload is publicly available for anyone to view at any time on the blockchain. Along with the transaction amount and timestamps, this payload reveals a concatenation of public keys. This comprises of the Bitcoin address and cryptographic signatures to provide an index linking the sender to the intended recipient of the Bitcoin (Nakamoto, 2008). Other analysis challenges exist as presented by cyber security researcher Kaminsky, 2011 in a 2011 Black Hat presentation on Bitcoin security when the Tor application is used. This application ensures anonymity *via* the Internet protocol stack leveraging the "Darknet" and utilizing a specific cryptocurrency "Dark Wallet" service. IP address obfuscation is achieved using a Tor router (Onion Router). IP address and Bitcoin address mappings are lost, and any investigator will only find the IP address associated to a Tor exit node preventing any meaningful analysis (Kaminsky, 2011).

Considering the limitations observed at the network layer when analyzing illicit Bitcoin activity, the next section reviews the literature relating to graph data models and how nodes and relationships formed on the Bitcoin network can provide insight into illicit activity.

GRAPH ANALYSIS

Directed Acyclic Graph (DAG)

A Directed Acyclic Graph (DAG)[1] is formed by the transactions and addresses on the Bitcoin network. The ability to break the entire Bitcoin graph into two smaller DAGs was researched by Reid and Harrigan (2011) as they investigated the problem of anonymity. A first DAG was constructed with Bitcoin addresses from tracing the flow of Bitcoins between users. A second DAG represented the analysis of transactions over time. The second DAG represented a transaction as a node and the directed edges between Bitcoin source and target were modeled as the output of one transaction to the input of another, creating a transaction chain. The graph may reveal transactions repeatedly performed by identifiable communities (multiple entities) or multiple transactions conducted by a single entity. Breaking the Bitcoin system down into two DAGs enables the ability to map and cluster behaviors of Bitcoin users and transactions over time. Reid and Harrigan (2011) break the Bitcoin system into analyzable user and transaction graphs and apply their method to reveal identity by using multiple sources of data. These data sources include: Off network information (building a directory of Bitcoin users) which allows monitoring activity, common transaction usage and routing behavior, using a website called the Bitcoin Faucet[2]. This website uses TCP/IP Network information, matching Bitcoin addresses to IP addresses, in order to build up a map of geographical usage. This could ultimately be flawed due to the Bitcoin propagation protocol where the last routed Bitcoin node IP address is not necessarily where the transaction originated. Examples of where the Bitcoin Faucet system has been applied include, looking at address pattern behavior attributed to known entities, such as WikiLeaks. In addition, using flow and temporal analyses to build a case study of Bitcoin theft.

Transaction Behavior

Taking algorithmic network analysis another step further helps the reader understand the evolutionary behavior of Bitcoin transactions and the way Bitcoin addresses adapt over time. Furthermore, advanced analytical techniques involving machine learning, can be used to determine the identity underneath the pseudonymous nature of Bitcoin addresses.

Ron and Shamir (2012), provide a step in this direction by analyzing a graph of the largest transactions in Bitcoin through a series of sub-graphs, identifying multiple characteristic behaviors for the flow of Bitcoin transactions. These are: "*long consecutive chains of transactions, fork-merge patterns that may include self loops, setting aside [Bitcoins] BTC's and final distribution of large sums via a binary tree-like structure.*" (Ron and Shamir, 2012). These patterns can be used to reflect common practice among users that may lead to suspicious behaviors on the Bitcoin network and these patterns can be re-used and applied to other illicit transaction scenarios. For example, Bartoletti et al. (2020) analyzed the redistribution of money flows relating to identifying Ponzi schemes in the cryptocurrency Ethereum. They identified several patterns in the money flows. The chain-shaped schemes and tree-shaped schemes are two illicit money-flow patterns that can also be modeled as a graph. To do this at any meaningful scale, automated software and algorithmic techniques are necessary. The following sections examine the literature relating to these techniques.

Automated Software

Spagnuolo et al. (2014) provide a framework for forensic analysis of such illicit Bitcoin transactions and subsequently developed graph analysis and automated software called Bitiodine. This software is used to parse the Bitcoin blockchain for transactions and addresses, and then augment that with different data scraped from the web to cluster, contextualize and visualize Bitcoin transaction graphs. An important piece to this literature is the application of their system to various case studies. These include investigating the Silk Road Bitcoin activity and associated trades made on the suspicious exchange, Mt Gox and transactions made by the owner of Silk Road, Dread Pirate Roberts, aka Ross Ulbricht, linking web forum data with blockchain data. Perhaps the most relevant application of Bitiodine is that of the Cryptolocker ransomware investigation. BitIodine is used "*to detect the CryptoLocker cluster(s), belonging to the malware authors, and compute some statistics about ransoms paid by the victims.*" (Spagnuolo et al., 2014). This data results from Google searches related to the ransomware, reddit forums that reveal addresses belonging to the ransomware, then a classifier is run over these addresses clustering the list of extorted addresses and automatically associating usernames from reddit to Bitcoin addresses. Furneaux (2018) also identifies several automated analysis tools that help visualize the Bitcoin graph and forensically investigate suspicious addresses. These tools include Numisight, Maltego, Learnmeabitcoin.com and the commercial enterprise systems available from Chainalysis and Elliptic which provide algorithmic modules to learn, infer and predict patterns in the network.

Algorithmic Analyses

Fleder et al. (2015) build on the previous techniques and look to identify suspicious behavior on the Bitcoin network. Providing context to the blockchain data from external data sources by web scraping forums and social media websites, graph analysis can be applied on the transactions performed to try and match any suspicious use of Bitcoin addresses. The methodology is similar to that found in Spagnuolo et al. (2014), however it introduces the use of the PageRank algorithm: "*We use PageRank as a guide to determine the most interesting nodes, or users in our user graph to further investigate their linkage with known forum users.*" (Fleder et al., 2015). The graph analysis techniques used (PageRank and clustering) are fundamental to a deeper behavioral analysis of the Bitcoin due to its inherent data structure, (the blockchain), and activity (transactions between users) forming a graph or network. According to Fleder et al. (2015), enriching the blockchain data by looking at external data in the form of security reports, Indicators of Compromise, malware sites and other cyber security feeds can help reveal identity for intelligence

[1]DAG – A directed edge (x, y) indicates that activity x must occur before y. They allow for topological sorting which is an important property providing order to process each vertex before any of its successors (Skiena, 2008).
[2]http://freebitcoins.appspot.com/

and law enforcement purposes. Particularly significant is the paper's use of the PageRank algorithm which is applied to the communities of transactions being performed by ransomware. This is a key indicator for understanding unusual behavior in networks, such as anomaly or fraud detection cases (Needham and Hodler, 2019).

As an example, Fleder et al. (2015) provided analysis on funds captured and sent to known Bitcoin addresses owned by the FBI. Nodes highly ranked *via* their technique were flagged for further investigation. Large clusters of transactions were detected from suspicious sites including WikiLeaks, cryptocurrency gaming service SatoshiDICE and the infamous Silk Road. The algorithmic technique from Fleder et al. (2015) borrows from other financial fraud risk management techniques. By associating an address or transaction coming from, or going to, such nefarious services as the Silk Road, it immediately becomes demarcated as a high-risk transactions or address on the Bitcoin network. Due to the potential risk of exposure to criminal activity a user has now made an illicit reference that can be tagged in the collected data. More advanced graph analysis techniques can be applied to sub-graphs of interest and reveal further intelligence on the Bitcoin network.

Although primarily concerned with the anonymity of Bitcoin, Gaihre et al. (2018) provide some important claims for analysis on transaction behavior, such as reuse frequency of addresses, zero balance addresses and how amounts are split up into smaller transactions with the usage of the change address, revisiting the concept of peeling introduced by Meiklejohn et al. (2013). Additionally, Gaihre et al. (2018), apply more advanced graph analysis techniques like the in-degree, which is the number of incoming edges to a node, as well as, connectedness of nodes on the network. Furthermore, they look at the diameter of the graph, which works on discovering the longest of all shortest paths in the network using a Bread First Search (BFS) algorithm. There are also several transaction walks that depict miner behaviors, where the miner accumulates the mined Bitcoin and also where the miner splits the mined Bitcoin. These can be useful payment typologies to build on for other illicit transaction activity.

Maesa et al. (2018) go deeper into the detail of the clustering algorithm used to generate a user graph containing nodes with groups of addresses controlling the transactions of interest. The clustering process is outlined step by step. This could be useful when applying a similar process to the population of incoming transactions to a ransomware seed address for example. This analysis yields a clustering coefficient of the user graph. A constant order of magnitude for the coefficient is exhibited over time and it is similar when compared to other complex social networks. Centrality measures provide the computation and interpretation of the results. These measures include PageRank and Eigenvector indexes to see the balance of nodes with respect to incoming and outgoing transactions. The Gini coefficient is also computed on the user graph, as a further measure to analyze the in-degree distribution over time. The Gini coefficient is an economic indicator that gauges economic inequality, measuring income distribution or wealth distribution among a population.

Another aspect Maesa et al. (2018) investigate is the analyses on the entire user graph of Bitcoin, as at the end of 2015. These analyses include a time series view of the Bitcoin network, along with economic analysis showing distribution of wealth. Furthermore, using techniques to detect critical nodes of the network where connectivity is strongest. The technique on node criticality is the most pertinent to illicit payment discovery. It is part of a centrality analysis on the graph and identifies the most active nodes in the graph. Nodes with high centrality (i.e., the most influential in a graph), will yield high in degree and/or out degree characteristics and Maesa et al. (2018) demonstrate a case that reveals the largest exchanges in the Bitcoin network, which at the time was Mt. Gox. This is also be applied to ransomware-Bitcoin analysis. For example, the centrality measures can reveal the most active nodes in a ransomware graph. Depending on the network depth, this could be the ransom seed address, the originating victim address (i.e., where the victim is getting their Bitcoin from), or the cash out point for where the cash out trail meets an exchange. This can become complex when interpreting whether the node actually has any influence over the movement of ransom payments during a ransomware campaign or simply over standard transactions on the Bitcoin network. That is why more information and context should be collected *via* machine learning to understand the representation of that node in the graph we are looking at.

MACHINE LEARNING TECHNIQUES

Machine learning in its simplest form is the act of teaching machines how to carry out tasks by themselves (Richert and Coelho, 2015). Richert and Coelho (2015) provide this introductory perspective in their book on building machine learning systems with python. The book provides a practical reference on building machine learning models in python to train a computer program to learn from data fed into a system. Richert and Coelho (2015) dive into the detail of the commonly used python programming language and the respective data science and statistical libraries needed to work through problem sets that required machine learning algorithm development as a solution to these problems. They highlight classification, topic modeling, sentiment analysis, regression, recommendation engines, computer vision and dimensionality reductions as important problem spaces to work on. The learning algorithms applied to these problems can take the form of supervised, unsupervised or reinforcement learning.

Kamath (2011) delivered a presentation at the annual python conference in 2011 that neatly summarized the differences in the available learning algorithms. Supervised learning is based on training data that contains correct responses to input data and as such the training data is used to learn a model that can be applied to classify future data items.

Unsupervised learning algorithms have no prior knowledge of the domain or structure of the data they use as inputs to interpret or classify meaningful outputs. It may not be possible to label the input data for the problem space being worked on, and unsupervised algorithms can be a powerful way to detect anomalies or learn features of the dataset being analyzed. One unsupervised learning method is clustering. This

is the process of grouping objects found in the input data exposing similar and distinctly different attributes which form clusters (Kamath, 2011). Bitcoin systems provide a strong case study for the clustering algorithm. An example of this can be realized with multiple input and multiple output Bitcoin transactions. Meiklejohn et al. (2013) found by grouping these types of transactions together it may be possible to find Bitcoin addresses and the transactions controlled by a common entity. Reinforcement learning provides a supervised and unsupervised hybrid learning approach. The learner runs through many different scenarios, then as a result of reinforcing an engineered policy against these scenarios, a good action is learned if it is a part of the well-engineered policy. Alpaydin (2020) comments on the goodness of policies, which is determined by a sequence of good actions which attain a desired goal.

Building on these learning techniques, the following literature looks at the analysis of Bitcoin networks using Machine Learning and Artificial Intelligence techniques with application to money laundering and fraud detection.

Supervised Machine Learning Techniques

Yin and Vatrapu (2017) analyze the clusters, entities and categories that are used to understand the control over funds in the Bitcoin network along with attributing some form of contextualization to the clusters with respect to the activity they are performing (e.g. Mining, mixing, exchanges). They also categorize based on criminal activity, in total the categories provided are Tor markets, scams, ransomware, mixing, and stolen bitcoins, exchange, gambling, merchant services, hosted wallets, mining pools, personal wallets. A methodology is provided outlining the data required from each cluster for analysis. This data includes: Transactions (hash, timestamp, input address, output address and value), addresses (address, number of transactions with peer address and value), counterparties (counterparty address, value, category and counterparty name), and exposure. Exposure acts as a risk calculation based on the knowledge of the cluster in terms of how many inputs and outputs out of total transactions emanate or arrive at a particular service category. The pipeline and analysis process diagram summarize the methodology having a big emphasis on data collection, cleansing, preparation and feature extraction. This reflects the high level of effort required to get the data ready to analyze. The second half of the diagram brings forth the machine learning capabilities for training data sets, model selection and validation. The statistical limitations on the machine learning components are identified in terms of the over and under sampling of the various classes, which limits the predictability of the under sampled classes. However, this methodology is something that can be refined with improved data collection, training and classification. This may be able to improve the 0.5 precision achieved on ransomware identification from their experiments.

Harlev et al. (2018) follow the same methodology as Yin and Vatrapu (2017) using supervised machine learning to attribute Bitcoin clusters to those predetermined categories. By looking at the anatomy of a Bitcoin cluster and using supervised machine learning to attribute Bitcoin clusters to those predetermined categories they break down the cluster structure to help categorize the controlling entities. Clustering will only take the analysis so far and emerging techniques based on neural networks that apply deep learning of latent representations on a graph or network structure provide an advantage. This is where the fraud team from Logical Clocks (2019) looked at the different machine learning approaches and how traditional AML anomaly detection problems use supervised machine learning against training data which contains an imbalance of "good" and "bad" transactions. They take this so far as saying it is an unviable approach which may only yield one bad transaction in more than a million. Therefore, there is a need to explore other machine learning methods to minimize the occurrence of the false positive and false negative detections and consequences of such detections.

Unsupervised Machine Learning Techniques

Whilst Yin and Vatrapu (2017) used supervised learning techniques, Monamo et al. (2016) provide a means of looking at the unsupervised learning techniques by giving the machine learning algorithms (trimmed k-means), which can both cluster objects and detect fraud in a multivariate setup to detect fraudulent Bitcoin activity. The k-means algorithm can perform clustering and classification without a training data set leaving the algorithm to establish its own labels as it comes across the data that is fed into it. This is both a limitation and a performance enhancement when it comes to fraud detection. Limitation in that unlabeled data somehow needs to be checked, modified and fed back into the system with context (manually). Performance enhancing as it will execute its machine components quicker. The authors concede that in the criminal detection process comparing known criminal elements would be better served using a neighborhood-based algorithm. These types of algorithms use classifiers to help the machine understand the context of the data they are processing and thus making the results more easily validated by experts in the field. Turner and Irwin (2018) experimented with the LINGO algorithm. They explain the open source nature of this algorithm and the previous application of the algorithm to web search results clustering by Osinski (2003). Osinski (2003) describes the algorithm as a combination of Latent Semantic Indexing (LSI) and the Vector Space Model (VSM) which use unsupervised and supervised learning techniques, respectively. The unsupervised application of LSI discovers abstract context in the data that passes through it. It forms cluster labels to be used as a reference for the supervised VSM algorithm. This is then used to determine cluster contents (Osinski, 2003). Turner and Irwin (2018) then look at applying LINGO to a combination of social media and Bitcoin blockchain data. Their results show a need to tune the algorithm with the input of subject matter expertise if any meaningful suspicious activity is to be found. Illicit money flows have traditionally been treated as anomaly detection problems. Researchers Graves and Clancy (2019) at DeepMind look to solve anomaly detection using unsupervised learning methods. One such advanced method seeks to train an algorithm to generate

its own models of the underlying classification of data it has discovered. These "generative" machine learning models can use common techniques such as k-means clustering and principal component analysis (PCA) to build a model of "good" and "illicit" transaction classes on the Bitcoin network. Such techniques can only be enabled through deep learning which provides a deep understanding of the data being observed in its context.

Deep Learning

Steenfatt et al. (2018) introduce an approach that allows deep learning on graph networks to learn the role a node plays in the network. This is based on the "struc2vec" algorithm, where traditionally similar nodes are in the same close proximity as each other, understanding the role a node plays with respect to embedded data yields node and network similarities that may not belong to directly connected components. Learning node representations or "node embeddings" that have meta-data and structural information encoded into them is a powerful way to find new suspicious relationships in the target network. An example given by Steenfatt et al. (2018) showed data from the WeChat payment network of 3,000 fraudulent nodes that have role labels from 15,000,000 nodes. The labels identified one of three types of fraud and grouped the transactions accordingly.

As an alternative to graph embedding, Li et al. (2019) proposed a Graph Matching Network (GMN), which calculates a graph similarity score by using Graph Neural Networks (GNN). GNNs are used to learn unlabeled graph structures by using the underlying encoded graph structured data (Zhang et al., 2009). Li et al. (2019) scale this idea up to work on complete graphs in order to understand their similarities by comparing the input graphs against different graphs to associate nodes and identify any differences in the node and edge features. This technique is related to the field of ransomware and through the application of graphs formed by ransomware–Bitcoin transactions the literature shows it is possible to understand the similarities and differences in a ransomware target network model. In addition, by creating a GNN for ransomware–Bitcoin graphs it is possible to machine train and learn what behaviors and parameters these networks may form in the future.

The collaboration between cryptocurrency forensic analysis firm Elliptic and researchers at IBM and Massachusetts Institute of Technology (MIT) have released a public data set of around 200,000 transactions partially labeled with illicit or non-illicit flags to identify suspicious transactions on the blockchain within the context of Anti-money Laundering (AML) (Weber et al., 2019). Using graph analysis techniques such as Graph Convolutional Networks (GCN) which use neural networks to allow the embedding of relational information between nodes and relationships to be further used in machine learning techniques. The GCN is a similar approach to the one taken by DeepWalk (Perozzi et al., 2014), however the difference is in the feature representations. A GCN aggregates the in and out degrees of a nodes neighbor and propagating these representations as features onto the nodes of the network. The DeepWalk embeds structural information on the graph to learn the typology of the graph by building up a node's context in the graph through a number of random walks from that node, much the same way a Natural Language Processing (NLP) algorithm learns words in a sentence from a corpus, or vocabulary, of words (Perozzi et al., 2014).

Furthermore, researchers at the CSIRO Data61 unit, produced a report on Bitcoin Ransomware Detection with scalable Graph Machine Learning (Jung, 2019). In this research, GCNs are also used to predict super nodes, those nodes in a Bitcoin network having a large amount of incoming and outgoing edges, which could be indicators of ransomware addresses and activity on the Bitcoin network.

Human and Machine

The techniques for examining the Bitcoin blockchain as a graph require a combination of machine powered analytics combined with human subject matter expertise in order to contextualize the data for intelligence collection and forensic interpretation. The ability to apply high performance computing to large amounts of data in the Bitcoin ecosystem provides efficiencies in analysis. Clustering data around influential nodes in the Bitcoin graph is a common approach undertaken by most of the authors of the literature. It allows for the application of graph algorithms relating to community detection, pageRank and centrality. Adding labels to the data collected and also combining the Bitcoin data with external data sources builds intelligence into the graph model by encoding structural knowledge into the graph such as in, out, or change addresses, timestamps, amount sent and received, service labels, network depth and address reuse frequency. A recent example of this is the open data project by Michalski et al. (2020) at the Harvard dataverse. They collected Bitcoin addresses and labeled them as mining pools, miners, coinjoin services, gambling services, exchanges, other services for training machine learning algorithms to learn and predict future addresses. A targeted application of these techniques is to the case of identifying ransomware payments in Bitcoin. At present there is limited application in this realm, however the intention is to look for similar graph patterns across different ransomware campaigns. Future research will be able to build upon these techniques and apply deep learning and Artificial Intelligence (AI) to further enhance the ransomware-Bitcoin target network model with labeled data and augment the cognitive process for identifying ransomware networks in the Bitcoin ecosystem.

RANSOMWARE–BITCOIN TRANSACTION ANALYSIS

Ransomware is a prevailing threat to the mainstream usage of cryptocurrencies and for malware developers and users, cryptocurrencies have enabled cyber criminals to collect their proceeds of crime undetected. Since 2018 the estimated global damage of ransomware has increased 2.5 times. From $US 8bn in 2018 to a projected $US 20bn in 2020 (Purplesec, 2020).

There is an essential need for identification and analysis frameworks. Ahn et al. (2016), describe a Ransomware Identification Framework (RIF) for identifying ransom payments from the set of all transactions sent to the ransomware cluster. Using cluster analysis on the total network of the Cryptolocker

ransomware campaign, they were able to understand the underlying financial infrastructures and money laundering strategies of the ransomware. Furthermore, the analysis yielded connections to popular services like BitcoinFog and BTC-e. It also speculated connections to criminal activity like the sheep marketplace, which was used for transacting narcotics, and was the successor to the infamous Silk Road site.

The methodology used by Ahn et al. (2016) for the RIF looks at the total number of transactions for each seed address, the total amount of bitcoins sent and received, and the number of ransom payments received. At an individual transaction level, the framework followed the input and output addresses, bitcoins transferred, and timestamps of these transfers. These parameters were used to build the target network model for their research, along with additional labels to indicate the network depth (i.e., how far away from the seed address the activity is taking place) and any service identifiers able to be picked up from a blockchain Application Programming Interface (API) that indicate Bitcoin exchanges.

Bistarelli et al. (2018) describe a tool that was created for this purpose. Through their analysis of the WannaCry attack, they were able to visualize the Bitcoin flows of WannaCry. Flows toward the three different ransom seed addresses were analyzed in an "in-flow" analysis to show a cluster of payments made to the ransom seed addresses and where they had come from. This revealed certain payments coming from leading crypto exchanges such as poloniex.com and other services like cubits.com. The "in-flow" analysis is one section of the intelligence-forensic continuum introduced as an analysis framework by Turner et al. (2019). It is important to take a full view of the continuum to build out the complete target network model, from mobilization through to actions on the objectives of the collected ransom.

Furthermore, Paquet-Clouston et al. (2018) analyze the collector addresses of the top 15 ransomware families by ransom payments received and by ransomware families. The authors investigate the graph formed by the incoming ransom payments and applied graph analysis techniques, such as centrality, to classify addresses to a particular ransomware. The two ransomware campaigns examined in detail from a graph analysis perspective were Locky and CryptoHitman. Transaction walks were produced showing which nodes in the graph acted as collectors and what services the addresses corresponded to, i.e., Bitcoin exchanges, mixing services, gambling services, etc. A longitudinal (time series) analysis was also conducted which showed the profile of a ransomware address and how it collected ransoms over time. Many of these profiles were similar, i.e., collecting their ransom over a burst of initial payments and then tapering off over the 1st week or two. Performing the time series analysis looks back at the history of a particular collector address and this is also important to understand the behavior of the victims and attacker. Paquet-Clouston et al. (2018), find that by moving back and forward through time over the lifespan of a Bitcoin address helps profile the incoming and outgoing relationships, providing a more targeted mechanism for identifying patterns in ransomware–bitcoin transaction graphs.

Patterns are one structure of interest providing a footprint to ransomware-Bitcoin activity. Another is measuring the impact or significance the ransomware attack had by plotting their collection and payment profiles. Conti et al. (2018) provide a "lightweight framework" to analyze 24 different types of ransomware from the perspective of their economic significance through the amount of Bitcoin they collected over time. The paper focuses solely on the number of Bitcoins received by the ransomware Bitcoin addresses over the time window for the ransomware campaign. They also look at the cumulative distribution function (CDF) of the ransomware to show the total amount of ransom collected over the campaign. This is a relatively simplified analysis that provides an approach to deal with some blockchain specifics on multiple input transactions and change addresses.

Huang et al. (2018) provide a more detailed insight into 10 ransomware clusters. The paper outlines a robust framework for identifying ransom addresses by scraping reports from real victims, creating synthetic victims under lab control conditions by making micropayments and tracing the flow of bitcoins and *via* clustering by co-spending which looks at addresses that create a transaction controlled by the ransom seed wallet. In addition, external data sources are looked at for information regarding the ransomware campaign. These include Google search history trends and YARA[3] malware indicators of compromise from a tool called VirusTotal. Once this framework has been set up and the initial detection and collection has been done, payment analysis can be conducted to look at things like estimating revenue of the ransomware, payment mechanics (timing and profile) and potential cash-out behavior. Cash-out behavior is one of the more interesting parts of the ransomware–bitcoin analysis as it gives targeted evidence on criminal behavior relating to ransomware attackers looking to use their proceeds of crime.

The techniques used for ransomware–Bitcoin analysis vary across the intelligence-forensics continuum using the elements discussed and by adding data attributes to nodes and vertices in a graph by labeling, it is possible to aid graph classification using graph machine learning algorithms to find similarity or trends in the graphs (Tiao et al., 2019). From the aforementioned literature, the importance of populating the target network model with context relevant data and comparing against different graphs from a variety of ransomware campaigns becomes evident.

DISCUSSION

The enforcement of AML/CTF KYC provisions for cryptocurrency will impede those who would misdirect its innovative functionality toward illicit ends and expose those who choose to do so. However, for law enforcement agencies to benefit, it is imperative that law enforcement agencies, financial

[3]YARA is a tool used in malware detection that creates rules based on hex, binary or string patterns that may be present as malware signatures in malicious files (Li, 2020).

intelligence units and cryptocurrency service providers should cooperate and share information. There is precedent for this.

For example, in 2017, a combined research and law enforcement partnership was made in the European Union between agencies and academic institutions from The Netherlands, Germany, Spain, Finland, Austria, and the UK, setting up the "Titanium" project, (Tools for Investigation of Transactions in Underground Markets). This project supported forensic analyses relating to criminal transactions, anomaly detection and machine learning techniques which were developed as a solution for investigations relating to criminal and terrorist acts using cryptocurrencies on the internet. According to Darknetmarkets (2017) Titanium was a platform using data from multiple sources, including "*online forums, P2P networks on dark marketplaces, virtual currencies and data found on electronic equipment that has been seized from suspects*." (Darknetmarkets, 2017). Demonstrating a strong partnership between technology and subject matter experts, Titanium is model project from which law enforcement can build upon to strengthen their role alongside technology in the discovery and fight against illicit cryptocurrency usage.

This paper reviewed various techniques that are quite limited on their own. However, in combination these techniques are a formidable arsenal, much greater than the sum of the individual techniques. These techniques range from the simple heuristic approaches that help assume ownership of addresses and transactions, to the graph algorithms that provide essential foundations for community detection, PageRank and connectedness patterns in illicit networks. Moreover, advanced computing power is enabling a resurgent field of Artificial Intelligence (AI). Machine Learning, when applied to graphs and networks, produces rich contextual understanding of graph behavior and opens new horizons for anomaly detection. It facilitates very detailed and complex benchmarking and pattern detection. Sophisticated algorithms such as Microcluster-Based Detector of Anomalies in Edge Streams (MIDAS), can detect dynamic behaviors in graphs (Mishra, 2018). This automated simultaneous analysis lends itself well to the Bitcoin–blockchain environment as the graphs formed here are constantly being updated with new addresses and transactions. This capability is particularly useful for ransomware attacks whose first indications are often sudden bursts of activity on the blockchain (Bhatia et al., 2019).

CONCLUSION

The literature reviewed in this paper forms a coherent approach to the analysis of the Bitcoin blockchain for illicit money flows. This approach revolves around techniques that seek to reduce the levels of anonymity provided by the Bitcoin system to identify real world participants. The literature reveals challenges with the regulatory environment. The different applications of laws and compliance controls across jurisdictions can hinder deanonymization and attribution to the real world of virtual identities on the cryptocurrency network. The emergence of machine learning and its application to graphs is providing a powerful analysis capability for disrupting Bitcoin related criminal activity. Particularly important are the practices of graph analysis, clustering, connectedness and GNNs as a form of deep learning applied to graphs. When compared to standard machine learning that employ supervised learning techniques and rules-based anomaly detection, these graph-based techniques dramatically enhance the future-orientated intelligence and real-time analysis of Bitcoin transactions.

Ultimately, the literature shows that there is no lack of available data on the Bitcoin blockchain. By providing open data this allows the community to flag certain behavior or orientation of Bitcoin addresses and transactions. However, the challenge is to correctly identify and classify the data and link it to off-chain data to provide a richer context. A way to potentially improve the performance of the machine learning algorithms is to take the graph labeling another step further. This would require adding more meta-data to the graph that attributes the addresses and transactions to various classifications, such as ransomware or other illicit purposes. These challenges have precipitated open data efforts such as those conducted by joint research collaborations at Harvard dataverse (Michalski et al., 2020) and between Elliptic, IBM and MIT (Weber et al., 2019) that will support future investigations and enhance intelligence sharing on illicit Bitcoin transactions.

AUTHOR CONTRIBUTIONS

AT: main author. SM and AU: corresponding authors, research supervisors, and editors. All authors contributed to the article and approved the submitted version.

REFERENCES

Ahn, G. J., Doupe, A., Zhao, Z., and Liao, K. (2016). "Ransomware and cryptocurrency: partners in crime," in *Cybercrime Through an Interdisciplinary Lens*, ed T. J. Holt (New York, NY: Routledge), 119–140.

Alpaydin, E. (2020). *Introduction to Machine Learning*, 4th *E*dn. Cambridge, MA: The Massachusetts Institute of Technology (MIT) Press.

Anti-Money Laundering and Counter-Terrorism Financing Act 2006. No. 169 (2006). *Compilation Date: 20 December 2018*. Available online at: https://www.legislation.gov.au/Details/C2019C00011/Html/Text (accessed July 22, 2020).

AUSTRAC (2019). *A Guide to Preparing and Implementing an AML/CTF Program for Your Digital Currency Exchange Business*. Available online at: https://www.austrac.gov.au/business/how-comply-guidance-and-resources/guidance-resources/guide-preparing-and-implementing-amlctf-program-your-digital-currency-exchange-business (accessed July 25, 2020).

Bartoletti, M., Carta, S., Cimoli, T., and Saia, R. (2020). Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. *Future Gener. Comput. Syst.* 102, 259–277. doi: 10.1016/j.future.2019.08.014

Bhatia, S., Hooi, B., Yoon, M., Shin, K., and Faloutsos, C. (2019). MIDAS: microcluster-based detector of anomalies in edge streams. *arXiv*. preprint arXiv:1911.04464. doi: 10.1609/aaai.v34i04.5724

Bistarelli, S., Parroccini, M., and Santini, F. (2018). "Visualizing Bitcoin flows of ransomware: WannaCry one week later", in *ITASEC, ser. CEUR Workshop Proceedings, no. 2058, 2018* [Online]. Available online at: http://ceur-ws.org/Vol-2058/#paper-13 (accessed November 6, 2018).

CipherTrace (2020). *Spring 2020 Cryptocurrency Crime and Anti-Money Laundering Report*. Available online at: https://ciphertrace.com/spring-2020-cryptocurrency-anti-money-laundering-report/ (accessed August 9, 2020).

Conti, M., Gangwal, A., and Ruj, S. (2018). On the economic significance of ransomware campaigns: a bitcoin transactions perspective. *Comput.* Security 79, 162–189. doi: 10.1016/j.cose.2018.08.008

Darknetmarkets (2017). *European Union Launches 'Titanium' Project Investigating Criminal Use of Bitcoin and Dark Web*. Available online at: https://darknetmarkets.co/european-union-launchestitanium-project-investigating-criminal-use-of-bitcoin-and-dark-web/ (accessed July 15, 2017).

Decree of the President of the Republic of Belarus No. 8. (2017). *National Legal Internet Portal of The Republic of Belarus, Dec. 27, 2017, No. 1/17415*. Available online at: http://pravo.by/document/?guid=12551andp0=Pd1700008andp1=1andp5=0 (in Russian), archived at https://perma.cc/7PJ7-YEKA, available in English at: http://law.by/document/?guid=3871andp0=Pd1700008e, archived at: https://perma.cc/V7UZ-TYM8 (accessed August 26, 2020).

Drainville, D. (2012). *An Analysis of the Bitcoin Electronic Cash System*. Waterloo, ON: University of Waterloo, 45.

Emsisoft (2020). *Report: The Cost of Ransomware in 2020. A Country-by-Country Analysis*. Emsisoft Malware Lab. Available online at: https://blog.emsisoft.com/en/35583/report-the-cost-of-ransomware-in-2020-a-country-by-country-analysis/ (accessed August 12, 2020).

EU (2018). *The 5th Anti-Money Laundering Directive*. Available online at: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-and-risk-management/anti-money-laundering-and-counter-terrorist-financing_en (accessed May 18, 2020).

FATF (2012). *The FATF International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation - the FATF Recommendations*. Paris. Available online at: https://www.fatf-gafi.org/publications/fatfrecommendations/documents/internationalstandardsoncombatingmoneylaunderingandthefinancingofterrorismproliferation-thefatfrecommendations.html (accessed July 25, 2020).

FATF (2019). *Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers*. Orlando, FL, United States. Available online at: https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets.html. (accessed July 25, 2020).

FIN-2019-G001 (2019). *Application of FinCEN's Regulations to Certain Business Models Involving Convertible Virtual Currencies*. Available online at: https://www.fincen.gov/sites/default/files/2019-05/FinCEN%20Guidance%20CVC%20FINAL%20508.pdf (accessed July 25, 2020).

FinCEN (2019). *New FinCEN Guidance Affirms Its Longstanding Regulatory Framework for Virtual Currencies and a New FinCEN Advisory Warns of Threats Posed by Virtual Currency Misuse*. Available online at: https://www.fincen.gov/news/news-releases/new-fincen-guidance-affirms-its-longstanding-regulatory-framework-virtual (accessed July 25, 2020).

Fleder, M., Kester, M. S., and Pillai, S. (2015). Bitcoin transaction graph analysis. *arXiv*. preprint arXiv:1502.01657.

Furneaux, N. (2018). *Investigating Cryptocurrencies: Understanding, Extracting, and Analyzing Blockchain Evidence*. Indianapolis, IN: John Wiley and Sons. doi: 10.1002/9781119549314

Gaihre, A., Luo, Y., and Liu, H. (2018). "Do bitcoin users really care about anonymity? an analysis of the bitcoin transaction graph," in *2018 IEEE International Conference on Big Data* (Seattle, WA: IEEE Big Data), 1198–1207. doi: 10.1109/BigData.2018.8622442

Graves, A., and Clancy, K. (2019). *Unsupervised Learning: The Curious Pupil*. Available online at: https://deepmind.com/blog/article/unsupervised-learning (accessed August 12, 2020).

Harlev, M. A., Sun Yin, H., Langenheldt, K. C., Mukkamala, R., and Vatrapu, R. (2018). "Breaking bad: De-anonymising entity types on the bitcoin blockchain using supervised machine learning," in *Proceedings of the 51st Hawaii International Conference on System Sciences* (Waikoloa Village, HI). doi: 10.24251/HICSS.2018.443

Huang, D. Y., McCoy, D., Aliapoulios, M. M., Li, V. G., Invernizzi, L., Bursztein, E., et al. (2018). "Tracking ransomware end-to-end," in *2018 IEEE Symposium on Security and Privacy (SP), 20–24 May 2018* (San Francisco, CA). doi: 10.1109/SP.2018.00047

Irwin, A. S., and Turner, A. B. (2018). Illicit Bitcoin transactions: challenges in getting to the who, what, when and where. *J. Money Laund. Control* 21, 297–313. doi: 10.1108/JMLC-07-2017-0031

Jung, K. (2019). *Bitcoin Ransomware Detection with scalable Graph Machine Learning*. CSIRO, Data61. Available online at: https://yowconference.com.au/slides/yowdata2019/KevinJung_BitcoinRansomwareDetection.pptx (accessed August 18, 2019).

Kamath, V. (2011). *Introduction to Machine Learning using Python*. Available online at: https://in.pycon.org/2011/static/files/talks/11/Introduction_To_ML_Partial_2.pdf (accessed March 12, 2016).

Kaminsky, D. (2011). *Some Thoughts on Bitcoin*. Available online at: http://dankaminsky.com/2011/08/05/bo2k11 (accessed July 15, 2015).

Karame, G. O., Androulaki, E., and Capkun, S. (2012). "Double-spending fast payments in bitcoin," in *Proceedings of the 2012 ACM Conference on Computer and communications security* (New York, NY), 906–917. doi: 10.1145/2382196.2382292

Li, V. (2020). *Intro to Malware Detection using YARA*. Available online at: https://medium.com/bugbountywriteup/intro-to-malware-detection-using-yara-eacab8373cf4 (accessed August 26, 2020).

Li, Y., Gu, C., Dullien, T., Vinyals, O., and Kohli, P. (2019). "Graph matching networks for learning the similarity of graph structured objects," in *Proceedings of the 36th International Conference on Machine Learning* (Long Beach, CA).

Logical Clocks (2019). *AI and Deep Learning for Fraud and AML*. Available online at: https://www.logicalclocks.com/blog/ai-deep-learning-for-anti-money-laundering (accessed August 11, 2020).

Maesa, D. D. F., Marino, A., and Ricci, L. (2018). Data-driven analysis of Bitcoin properties: exploiting the users graph. *Int. J. Data Sci. Anal.* 6, 63–80. doi: 10.1007/s41060-017-0074-x

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., et al. (2013). "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 Conference on Internet Measurement Conference* (New York, NY), 127–140. doi: 10.1145/2504730.2504747

Michalski, R., Macek, P., and Dziubałtowska, D. (2020). *Bitcoin Addresses and Their Categories*, Harvard Dataverse, V1, UNF:6:yMWT5M5ZVtmYA+B5iUqf3Q== [fileUNF]

Mishra, N. (2018). *Anomaly Detection in Dynamic Graphs Using MIDAS: An Interesting Approach to Modelling Network Security*. Available online at: https://towardsdatascience.com/anomaly-detection-in-dynamic-graphs-using-midas-e4f8d0b1db45 (accessed May 18, 2020).

Monamo, P., Marivate, V., and Twala, B. (2016). "*Unsupervised Learning for Robust Bitcoin Fraud Detection*," in *2016 Information Security for South Africa (ISSA)* (Johannesburg: IEEE), 129–134. doi: 10.1109/ISSA.2016.7802939

Nakamoto, S. (2008). *A Peer-to-Peer Electronic Cash System*. Bitcoin. Available online at: https://bitcoin.org/bitcoin.pdf (accessed August 13, 2020).

Needham, M., and Hodler, A. E. (2019). *Graph Algorithms: Practical Examples in Apache Spark and Neo4j*. Sebastopol, CA:O'Reilly Media, Inc.

Osinski, S. (2003). *An algorithm for clustering of web search results* (Master thesis). Poznan: Poznan University of Technology.

Paquet-Clouston, M., Haslhofer, B., and Dupont, B. (2018). "Ransomware payments in the bitcoin ecosystem," in *The 17th Annual Workshop on the Economics of Information Security (WEIS)* (Innsbruck). doi: 10.1093/cybsec/tyz003

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). "Deepwalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY), 701–710. doi: 10.1145/2623330.2623732

Pilarowski, G., and Yue, L. (2017). *PBOC, CAC, MIIT, SAIC, CBRC, CSRC, and CIRC, Announcement on Preventing Financial Risks from Initial Coin Offerings*. Available online at: http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3374222/index.html (in Chinese), archived at: https://perma.cc/N88N-5CV5. *See Greg Pilarowski and Lu Yue, China Bans Initial Coin Offerings and Cryptocurrency Trading Platforms, CHINA REGULATION WATCH*. Available online at: http://www.pillarlegalpc.com/en/news/2017/09/21/china-bans-initial-coin-offerings-and-cryptocurrency-tradingplatforms/, archived at: https://perma.cc/VQ2W-T4HY (accessed August 28, 2020).

Purplesec (2020). *The Growing Threat Of Ransomware*. Available online at: https://purplesec.us/resources/cyber-security-statistics/ransomware/ (accessed August 12, 2020).

Reid, F., and Harrigan, M. (2011). "An analysis of anonymity in the bitcoin system," in *2011 International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing* (Boston, MA). doi: 10.1109/PASSAT/SocialCom.2011.79

Richert, W., and Coelho, L. P. (2015). *Building Machine Learning Systems With Python, 2nd Edn, Get More from Your Data through Creating Practical Machine Learning Systems with Python*. Birmingham: Packt Publishing.

Ron, D., and Shamir, A. (2012). *Quantitative Analysis of the Full Bitcoin Transaction Graph*. The International Association for Cryptologic Research (IACR) Cryptology ePrint Archive, 584.

Rosenfeld, M. (2011). Analysis of bitcoin pooled mining reward systems. *arXiv*. preprint arXiv:1112.4980.

Skiena, S. (2008). *The Algorithm Design Manual*. London: Springer-Verlag London. doi: 10.1007/978-1-84800-070-4

Spagnuolo, M., Maggi, F., and Zanero, S. (2014). "Bitiodine: extracting intelligence from the bitcoin network," in *International Conference on Financial Cryptography and Data Security* (Berlin, Heidelberg: Springer), 457–468. doi: 10.1007/978-3-662-45472-5_29

Steenfatt, N., Nikolentzos, G., Vazirgiannis, M., and Zhao, Q. (2018). "Learning structural node representations on directed graphs," in *International Conference on Complex Networks and their Applications* (Cham: Springer), 132–144. doi: 10.1007/978-3-030-05414-4_11

Stokes, R. (2012). Virtual money laundering: the case of Bitcoin and the Linden dollar. *Inf. Commun. Technol. Law* 21, 221–236. doi: 10.1080/13600834.2012.744225

The Law Library of Congress (2018). *Regulation of Cryptocurrency in Selected Jurisdictions*. Available online at: https://www.loc.gov/law/help/legal-reports.php; https://www.loc.gov/law/help/cryptocurrency/regulation-of-cryptocurrency.pdf (accessed July 25, 2020).

Tiao, L., Elinas, P., Nguyen, H., and Bonilla, E. V. (2019). Variational spectral graph convolutional networks. *arXiv*. preprint arXiv:1906.01852.

Tsukerman, M. (2015). The block is hot: a survey of the state of Bitcoin regulation and suggestions for the future. *Berkeley Technol. Law J.* 30, 1127–1170. Available online at: https://www.jstor.org/stable/26377750

Tu, K. V., and Meredith, M. W. (2015). Rethinking virtual currency regulation in the Bitcoin age. *Wash. L. Rev.* 90, 271–347.

Turner, A., and Irwin, A. S. M. (2018). Bitcoin transactions: a digital discovery of illicit activity on the blockchain. *J. Finan. Crime* 25, 109–130. doi: 10.1108/JFC-12-2016-0078

Turner, A., McCombie, S., and Uhlmann, A. (2019). A target-centric intelligence approach to WannaCry 2.0. *J. Money Laund. Control* 22, 646–665. doi: 10.1108/JMLC-01-2019-0005

Wagstaff, J., and Karpeles, M. (2014). *Mt. Gox Bitcoin Debacle: Huge Heist or Sloppy Glitch*. Reuters. Available online at: https://www.reuters.com/assets/print?aid=USBREA1R0Y720140228 (accessed July 22, 2020).

Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., et al. (2019). "Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics," in *Tutorial in the Anomaly Detection in Finance Workshop at the 25th SIGKDD Conference on Knowledge Discovery and Data Mining* (Anchorage, AK).

Yin, H. S., and Vatrapu, R. (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* (Boston, MA: IEEE, Big Data), 3690–3699.

Zhang, S. J., Hagenbuchner, M., Scarselli, F., and Tsoi, A. C. (2009). "Supervised encoding of graph-of-graphs for classification and regression problems," in *International Workshop of the Initiative for the Evaluation of XML Retrieval* (Berlin, Heidelberg: Springer), 449–461. doi: 10.1007/978-3-642-14556-8_45

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.