

Detecting Anomalous Cryptocurrency Transactions: an AML/CFT Application of Machine Learning-based Forensics

Nadia Pocher^{1,3*}, Mirko Zichichi^{2,3}, Fabio Merizzi³, Muhammad Zohaib Shafiq³ and Stefano Ferretti⁴

¹ Institute of Law and Technology, Faculty of Law, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain .

² Ontology Engineering Group, Universidad Politécnica de Madrid, ETSIINF, Boadilla del Monte (MD), 28660, Spain .

³ Department of Computer Science and Engineering and Department of Legal Studies, University of Bologna, Bologna, 40126, Italy .

⁴ Dipartimento di Scienze Pure e Applicate, University of Urbino “Carlo Bo”, Piazza della Repubblica, 13, Urbino, 61029, Italy .

*Corresponding author(s). E-mail(s): nadia.pocher@uab.cat;

Contributing authors: mirko.zichichi@upm.es;

fabio.merizzi@studio.unibo.it;

muhhammad.shafiq6@studio.unibo.it; stefano.ferretti@uniurb.it;

Abstract

The rise of blockchain and distributed ledger technologies (DLTs) in the financial sector has generated a socio-economic shift that triggered legal concerns and regulatory initiatives. While the anonymity of DLTs may safeguard the right to privacy, data protection and other civil liberties, lack of identification hinders accountability, investigation and enforcement. The resulting challenges extend to the rules to combat money laundering and the financing of terrorism and proliferation (AML/CFT). As law enforcement agencies and analytics companies have begun to successfully apply forensics to track currency across blockchain ecosystems, in this paper we focus on the increasing relevance of these techniques. In particular, we offer insights into the application to the Internet of

2 *Detecting Anomalous Cryptocurrency Transactions*

Money (IoM) of machine learning, network and transaction graph analysis. After providing some background on the notion of anonymity in the IoM and on the interplay between AML/CFT and blockchain forensics, we focus on anomaly detection approaches leading to our experiments. Namely, we analyzed a real-world dataset of Bitcoin transactions represented as a directed graph network through various machine learning techniques. Our claim is that the AML/CFT domain could benefit from novel graph analysis methods in machine learning. Indeed, our findings show that the Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) neural network types represent a promising solution for AML/CFT compliance.

Keywords: blockchain technology, financial technology, network forensics, graph analysis, AML/CFT

1 Introduction

Over the last 15 years, the application of blockchain and distributed ledger technologies (DLTs) to the financial domain has been surrounded by an enthusiastic hype. Building on years of research in distributed systems and cryptography, the launch of Bitcoin (Nakamoto, 2008) finally allowed to record information in the form of transactions and agree on their order without the need to rely on a trusted third-party. This revolution opened the way to peer-to-peer value transfers, thus enabling citizens and businesses to participate directly in a new digital global economy. The advent of a token-based “Internet of Value” (IoV) (Tapscott & Euchner, 2019) and of the “Internet of Money” (IoM) (Antonopoulos, 2017) led to a socio-economic transformation that triggered legal concerns and spurred regulatory initiatives. Chiefly, disintermediation and the perceived anonymity of these transactions vis-à-vis blockchain’s transparency defied schemes of accountability and generated fears of exploitation for large-scale illicit purposes. Indeed, challenges arose as to the application of the rules to combat money laundering and the financing of terrorism and proliferation (AML/CFT)¹, where identification is a crucial concept, *e.g.*, know your customer.

The AML framework consists of a set of laws, regulations and procedures, preventive and repressive, that aim to protect the integrity of the financial system by hampering the concealment of the origin of illicit profits to hinder their enjoyment (Authors blinded for review, n.d.). To this end, compliance duties are thrust on specific entities, ranging from financial institutions to professionals (*e.g.*, lawyers, notaries) to art galleries. The international AML action is overseen by the Financial Action Task Force (FATF), an intergovernmental, policy making, monitoring, enforcement and standard-setting body that provides guidance through its Recommendations (FATF, 2022). Although they

¹For brevity, in the remaining paper AML refers to both AML/CFT.

are soft-law instruments, not directly binding on individuals and organizations, FATF’s members committed to transposing them into domestic law. In the EU, they are implemented through a series of “AML Directives”, while a recent legislative initiative dubbed “AML Package” proposes the establishment of a EU-wide single rulebook on AML ([European Commission, 2021](#)).

In this context, the link between cryptocurrencies and anonymity was emphasized not only by their (debated) cyberlibertarian imprint, but also by scandals that reached the headlines in the first years of the IoM, including the shutdown of darknet markets such as the Silk Road in 2013 and AlphaBay and Hansa Market in 2017. In facilitating the exchange of illicit products and services, these platforms were leveraging Bitcoin’s (perceived) anonymity to avoid regulation and law enforcement ([Goforth, 2020](#)). The related risk perception was confirmed by several exchange hacking incidents – *e.g.*, Mt Gox in 2014, later Bitfinex, Coincheck and Bitgrail, with a combined loss of more than USD 1.2 billion ([Johnstone, 2021](#)). Meanwhile, in 2021 crypto-related laundering amounted to USD 8.6 billion, with a value of at least USD 10 billion currently held by illicit addresses, while it appears heavily concentrated: most value originating from illicit addresses is seemingly sent to few services, often built for criminal purposes ([Chainalysis Team, 2022](#)).

Thus, the picture of the IoM as a realm of untraceable transfers and individual freedom from governmental control warrants a two-fold interpretation: while anonymity can safeguard the right to privacy, data protection and other civil liberties, lack of identification hampers investigation, enforcement and accountability. Two sets of mutually influencing events emerged in this context. On the one hand, investigative authorities and law enforcement agencies started developing forensic techniques to “follow the money” across blockchain ecosystems ([Bartoletti, Carta, Cimoli, & Saia, 2020](#); [Biryukov & Tikhomirov, 2019](#); [Chen, Zheng, Ngai, Zheng, & Zhou, 2019](#); [Lischke & Fabian, 2016](#); [Meiklejohn et al., 2016](#); [Moreno-Sanchez, Zafar, & Kate, 2016](#); [Neudecker & Hartenstein, 2017](#); [Phan, 2021](#); [Yin, Langenheldt, Harlev, Mukkamala, & Vatrapi, 2019](#)). On the other hand, the unveiled insufficiency in Bitcoin’s anonymity spurred altcoin projects such as Monero and ZCash to implement an advanced set of cryptographic methods that require a variety of new advanced analytical tools.

Against this backdrop, in this paper we focus on the increasing value of intelligence techniques to provide insights into IoM ecosystems, with specific regard to techniques of machine learning, network and transaction graph analysis ([Fleder, Kester, & Pillai, 2015](#); [Ober, Katzenbeisser, & Hamacher, 2013](#); [Weber et al., 2019](#); [Y. Wu et al., 2021](#)). Notably, we first provide some background on the notion of anonymity in the IoM landscape and on the interplay between AML and blockchain forensics. Consequently, we focus on the anomaly detection approaches that led to the experiments that we performed on a real world transactions dataset. In particular, a real Bitcoin transactions dataset represented as a directed graph network was analyzed through several machine learning techniques. Our main hypothesis was that, since money laundering

4 *Detecting Anomalous Cryptocurrency Transactions*

involves transactions flow relationships between entities creating a graph structure, then AML analytics could benefit from novel graph analysis techniques in machine learning, namely Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). Moreover, in developing this work we heed the following assumptions:

- while techniques of cryptocurrency forensics are manifold, our work does not aim to review them comprehensively;
- cryptocurrency transactional data is often analyzed through a combination of on-chain and off-chain forensic techniques, thus including information not recorded on blockchain or recorded on a different blockchain; however, in this work we focus on on-chain data;
- the IoM is neither a legal nor a technical definition; in this paper, the term broadly refers to the entire set of cryptocurrency ecosystems, thus including the part of the Internet of Value that relates to payment tokens;
- research into explainability and interpretability of artificial intelligence and machine learning applications, as well as relevant legal impacts, is vast and detailed; in light of the scope of this work, we perform inevitable simplifications.

The results of our experiments show that GCN and GAT neural network typologies represents a promising solution for AML. This is in opposition to a state of the art work in which a baseline supervised learning algorithm, *i.e.*, non-graph-based, such as the Random Forest provided the best performances (Weber et al., 2019). In this context, we underline the value of experimenting with techniques based on machine learning and transaction graph analysis, and with their possible combinations. We contextualize our argument considering the significant amount and complexity of transaction data to be processed in the IoM, and the specifics of the rules-based anomaly indicators provided by the AML framework. We do so by taking into account the need to mitigate the shortcomings of rules-based regimes, explainability aspects and the need to engage into research that deploys an interdisciplinary methodology.

The remainder is structured as follows. Section II provides a conceptual background on Bitcoin’s pseudonymity, and insights into the relationship between forensic techniques and the AML regulatory framework. Section III explores related work. Section IV takes a context-specific approach to outline anomaly detection techniques. In Sections V and VI, respectively, we present and discuss our study on the use of machine learning-based classification methods for AML purposes. Section VI concludes the paper.

2 Background

2.1 Pseudonymity and De-Anonymization

While untraceable payments are a perfect fit for cyberlibertarian dreams (Filippi & Wright, 2018), the IoM is populated by manifold socio-technical

notions of anonymity and related concepts such as transparency, privacy and auditability (Authors blinded for review, n.d.). Surveying the relevant research (Amarasinghe, Boyen, & McKague, 2019) falls outside our work’s scope. Thus, we heed a domain-specific understanding that anonymity in the IoM “means being able to conduct a financial transaction without anyone besides the sender and the receiver being able to identify the parties involved” (Edmunds, 2020). In this respect, it is a common blockchain goal to combine user anonymity and transparency of operations, and a public blockchain is structurally designed to enable anonymous peer-to-peer transfers (Quiniou, 2019). However, blockchain types feature different approaches to identification: in public permissionless ledgers, with no centralized authority, nodes operate with no association to a specific identity, while in permissioned ledgers a centralized entity/consortium identifies nodes whose key-pair is usually associated with a real-world identity (Wang & De Filippi, 2020)

The definition of Bitcoin as *anonymous* has been challenged extensively and it is widely agreed that the correct label for it, as well as for other cryptocurrencies, is *pseudonymous* (Berg, 2019; Biryukov & Tikhomirov, 2019; Y. Li et al., 2019). *Pseudonymity* refers to the use of pseudonyms as identifiers, and a *pseudonym* is a subject’s identifier other than the subject’s real names (Pfitzmann & Hansen, 2010). In most blockchain systems, public-private key pairs are the identifiers that uniquely identify the holder of the wallet (Wang & De Filippi, 2020). From this perspective, in a cryptocurrency transaction, addresses (*i.e.*, public keys) perform a “username” function. Hence, senders and recipients are pseudonymous, and not anonymous, when they are identified by their address. In a public ledger such as Bitcoin’s, the transaction history is transparent but the network’s participants are linked to addresses, not to specific identities (Amarasinghe et al., 2019). In particular, a transaction references previous transaction “outputs” as new transaction “inputs” and it is validated whenever the address indicated in the previous outputs digitally signs the new transaction, *i.e.*, demonstrate the ownership of those Bitcoin values.

In principle, a currency scheme aims to avoid that the transaction history of a specific unit can be retraced. If it is technically possible to associate a coin with its past exchanges, the currency’s fungibility is threatened and the nominal value of the given unit is affected. Insofar as unit traceability generates a credibility loss, the property of anonymity emerges as paramount. Because Bitcoin’s features seemed insufficient, new techniques have been embedded into anonymity enhanced currencies (AECs), also known as privacy coins, leveraging Privacy Enhancing Technologies (PETs) and notably Zero-Knowledge Proofs. Besides fungibility, AECs’ developers aim for user unaccountability and privacy, striving to bypass regulatory constraints but also governmental surveillance. The state of privacy pursued by privacy coins, however, is almost always anonymity, the pillar of most AEC whitepapers (Harvey & Branco-Illodo, 2020). The advent of complex anonymity-oriented models was

accompanied by the increasing interest of institutional and corporate stakeholders in the IoM, with stablecoin projects and central bank digital currencies (Authors blinded for review, n.d.) . In this context, the array of privacy and data protection concerns grew wider as to the possible exploitation of financial data by a broader range of actors.

As Bitcoin became popular and privacy coins were developed, experts and law enforcement professionals devised strategies to trace cryptocurrency transfers. In this context, the end goal of intelligence methods is to match users, definitively or statistically, to transactions performed by crypto-addresses – *i.e.*, to connect pseudonyms to real-world identities – leveraging unique identifiers. These techniques started to be deployed under the label of “blockchain forensics”, as they were informed by the specificities of blockchain technology. Accordingly, they were defined as the use of science and technology for the sake of investigation and fact-establishment in a court of law, primarily dealing with recovering and analyzing evidence on blockchain ledgers resulting from transaction activities (Phan, 2021). Later, analytic solutions started to be requested by regulated entities. Although they have been mostly tested on the Bitcoin network, data-exploitation strategies have been deployed on the Ethereum blockchain (Bartoletti et al., 2020; Chen et al., 2019; Y. Li et al., 2021; Moreno-Sanchez et al., 2016) , and on non-blockchain DLTs (Ince, Liu, & Zhang, 2018; Tennant, 2017).

Since identifiers (*i.e.*, addresses/public keys) can be leveraged to connect transactions to their history, Bitcoin’s pseudonymity generates an inherent tension between anonymity and accountability (Yin et al., 2019). Unless they are associated with additional data, however, identifiers do not reveal personal identifying information (Wang & De Filippi, 2020). Hence, pseudonymity does not imply identifiability, which is subjective: a pseudonymous subject is identifiable only if for a specific actor it is possible to discover its real-world identity. This concept is crucial in the IoM, as two forces oppose each other: there are actors, such as authorities and exchange/forensic service providers, that seek to achieve identification, while many strategies are employed at various levels to avert it, *e.g.*, advanced cryptography and virtual private networks. Indeed, technology can both foster new pathways to accountability and disrupt data retrievability. In the IoM, anonymity enhancements are linked to forensics, which influences the overall anonymity level vis-à-vis users’ strategies to limit fund traceability. An example of these anonymity-enhancing methods is the use of “self-hosted/unhosted” wallets, also known as “non-custodial”, where users hold their private keys directly as opposed to “hosted/custodial” wallets where storage and custody are offered as a service by a third party (Authors blinded for review, n.d.) , usually an AML regulated entity. In this context, the transparent nature of (public) blockchains makes them vulnerable to insufficient data privacy, de-anonymization attacks and possible application of surveillance techniques. While de-anonymization is often perceived negatively, it can be applied for investigative purposes and to comply with rules that aim to mitigate specific risks.

2.2 AML/CFT and Blockchain Forensics

The first concerns of cryptocurrency misuse originated from events where their (purported) anonymity was linked to illicit transactions on the dark web. While a range of technologies aid darknet operations, cryptocurrencies, mostly Bitcoin and Monero, play a key role by facilitating payments (Akhgar, Gercke, Vrochidis, & Gibson, 2021). While Bitcoin is still the major player, used by 93% of darknet markets, the adoption of Monero is increasing: 67% of platforms supported it in 2021 vis-à-vis 45% in 2020, and some support it on an exclusive basis (Chainalysis Team, 2022). Nonetheless, forensic experts argue the public perception of cryptocurrency-related laundering is inflated. Indeed, between 2019 and 2020 crypto-crime seemingly decreased by 57%, from USD 4.5 billions to 1.9 billions, with criminal activities being 160 times more likely to involve fiat currencies (CipherTrace, 2021; Goforth, 2020). Accordingly, even if the value of illicit crypto-transactions reached an all-time high in 2021, up 79% from 2020, data show an all-time low in terms of crypto-activity share. Indeed, the total transaction volume grew by 567% between 2020 and 2021, reaching USD 15.8 trillion, with criminal usage representing only 0.15%.

Despite the decrease in their number, in 2021 darknet markets' revenue set a record of USD 2.1 billion (Chainalysis Team, 2022). Since AML obligations are informed by the risk-based approach, regulated entities must tune compliance efforts to the principle of proportionality: stricter measures if risk factors are higher, and consistent internal procedures. They must consider risk factors predefined by the FATF, the EU, national law, sector-specific supervisors, known as "risk/red flag/anomaly indicators". The end-goal is to draw the authorities' attention when suspicions of illicit activities arise. The submission of suspicious transaction reports is the core duty that orients other measures such as licensing regimes, customer due diligence obligations such as know your customer and ongoing monitoring (*e.g.*, transaction scrutiny). A suspicious transaction report must be submitted when the entity knows, suspects, or has reasonable ground to suspect, the given funds are the proceeds of a criminal activity, or are related to terrorist financing (Directive (EU) 2018/843, 2018; FATF, 2022).

Generally, AML duties apply to crypto-transactions, and a set of related service providers are regulated. In the EU, the 5th AML Directive (Directive (EU) 2018/843, 2018) first targeted these activities, and the regime is evolving with the AML Package (European Commission, 2021). Even if some blockchain features mitigate the risk of fraudulent behaviour, anomalies are not detected automatically and the technology is vulnerable to unpredictable exploitation methods (Shayegan, Sabor, Uddin, & Chen, 2022; Xu, 2016), which drove the development of specific techniques of anomaly detection. In this respect, the IoM's opaque reputation appears paradoxical, since it provides a huge amount of open-source intelligence – *e.g.*, it is possible to extract data from a given transaction and retrieve the history of an address, while methods using networks created by transactions (*i.e.*, "transaction flow analysis") can define

patterns to pinpoint suspected addresses (Y. Wu et al., 2021). Different analytic techniques have been refined over time (Yin et al., 2019), and mostly rely on statistical approaches – *e.g.*, the re-use of the same account for more transactions or the co-use of more accounts for a single transaction can lead to match more accounts to the same user (Y. Li et al., 2021).

In 2020 a surge of ransomware attacks showed framework deficiencies vis-à-vis the complexity of the global development of the IoM, increasingly populated by privacy coins, mixers, decentralized platforms, unhosted wallets, other products and services enabling and/or allowing reduced transparency and increased obfuscation, and also innovative models like initial coin offerings (Custers & Overwater, 2019), featuring fraud and money laundering risks. Accordingly, in September 2020 the FATF published virtual-asset-specific red flag indicators, with a section on anonymity risks (FATF, 2020).² From this perspective, it was argued AML provisions display an understanding of anonymity that does not distinguish between “anonymization” and “strong pseudonymization”, and refers to both the (near) impossibility of linking data on a ledger with (an) identified person(s), and a situation in which the linking capacity is significantly hampered (Karasek-wojciechowicz, 2021).

Although a transaction’s anonymity level is not sufficient to suggest the transfer is suspicious, the FATF underlined inherent issues of PETs implemented by privacy coins, such as Zero-Knowledge Proofs (FATF, 2020), while a range of institutions highlighted the risks caused by privacy-enhanced/unhosted wallets (Europol, 2020). Against this backdrop, the application of forensic techniques appears pivotal at different levels: law enforcement agencies and authorities deploy them for investigative and enforcement purposes, and analytic companies develop RegTech³ solutions for regulated entities to comply with AML requirements such as risk assessments, identification, traceability. Overall, intelligence methods provide a wide range of information on IoM ecosystems that can be used to various ends. Notably, they are crucial when confronting the technological problems encountered by service providers to comply with the debated “crypto travel rule”, pursuant to which they have to ensure identifiability of originators and recipients of crypto-transfers.

The term “crypto travel rule” dubs the expansion to the IoM of information sharing measures previously applicable only to wire transfers, as per FATF Recommendation 16 and Regulation (EU) 2015/847 (Fund Transfers Regulation). These rules require financial institutions to ensure traceability throughout the payment chain. Accordingly to an update of the FATF Standards, the AML Package (European Commission, 2021) proposes to recast the Fund Transfers Regulation, expanding its scope and adjusting it to cryptocurrency features. Although the application of AML rules to the IoM has long shown its weaknesses, the “crypto travel rule” exacerbates the tension

²The report targets six types of indicators, relating to (i) transactions, (ii) transaction patterns, (iii) anonymity, (iv) senders/recipients, (v) funding/wealth at source, (vi) geographic risks.

³The term “RegTech”, short for “regulatory technology”, refers to the use of new technologies to aid regulatory and compliance processes, mostly through FinTech software applications.

between the evolution of these ecosystems and an intermediary-based framework. For instance, ongoing debates concern transfers between self-hosted and hosted wallets, and the challenge of linking IoM activity to real-world identities (identification of originators and beneficiaries) (Authors blinded for review, n.d.) . Meanwhile, the industry denounces the absence of global standards and technical solutions to underpin effective and affordable compliance.

3 Related Work

In this section we will describe some works that provide the application of many concepts introduced in the previous section regarding blockchain de-anonymization and forensics.

In blockchain analytics, various methods aim to link (pools of) addresses and transactions to specific users and rely on the concepts of “address clusters” (Ince et al., 2018; Neudecker & Hartenstein, 2017) and “transaction graphs” (Al Jawaheri, Al Sabah, Boshmaf, & Erbad, 2020; Fleder et al., 2015; Ober et al., 2013; Weber et al., 2019). They usually involve visualization analytics and pursue the clustering of addresses owned by the same user (Y. Wu et al., 2021), the identification of “idioms of use” in the network that can erode anonymity (Meiklejohn et al., 2016), while some approaches screen transactions to/from crypto-wallets to classify each transaction as licit or illicit (Weber et al., 2019). These tools do not (try to) de-anonymize addresses and transactions, linking them to real-world identities, but in case one of them is de-anonymized (in other ways) they allow to de-anonymize the whole cluster, as the cluster database allows fast correlation. Likewise, the goal is normally not to identify and analyze transaction patterns, but to allow that, if a cluster’s address is suspected, other addresses likely owned by the same user/group can be suspected as well (Y. Wu et al., 2021).

These methodologies are based on “clustering heuristics” models (Androulaki, Karame, Roeschlin, Scherer, & Capkun, 2013; Lischke & Fabian, 2016; Meiklejohn et al., 2016; Reid & Harrigan, 2013), such as: if two/-more addresses are inputs to the same transaction, they are controlled by the same user (Meiklejohn et al., 2016). The heuristics are applied by “wallet-closure analysis” to establish a unique many-one mapping between addresses and an identity (Al Jawaheri et al., 2020), which ultimately allows to infer several links between the user and hidden services to be identified.

Relatedly, “behaviour-based clustering” (Yin et al., 2019) groups addresses based on patterns such as transaction values (Amarasinghe et al., 2019). A study performed by (Androulaki et al., 2013) showed this technique can unveil the profiles of 40% of Bitcoin users despite privacy measures. On the application level, analytic techniques can exploit information that leaks when it is possible to establish correlations between transactions and users’ profiles on social networks. Frequently users post their addresses (*e.g.*, to receive donations, offer services) but also reveal personal information (*e.g.*, contact information, age, gender, location) (Al Jawaheri et al., 2020). In this respect,

“transaction fingerprinting” methods can leverage “publicly available” or “off-network” information (Reid & Harrigan, 2013), which is also the context of application of techniques involving web-scraping and Open Source Intelligence tools. The authors in (Fleder et al., 2015) annotated the transaction graph by linking user pseudonyms to online identities collected from social networks, and developed a graph-analysis framework to summarize and cluster users’ activity to link identities and transactions.

Specific methods target “mixing” services (J. Wu et al., 2020), *i.e.*, the ones that shuffle cryptocurrency funds with others by sending them to different addresses, to obfuscate a flow that would otherwise allow to retrace the original owner. Although third-party services act as centralization points for traceability purposes, new disintermediated methods, such as the peer-to-peer protocol CoinJoin (Al Jawaheri et al., 2020), pursue shuffling goals through more sophisticated approaches. In this context, an important role is played by peer-to-peer transactions between cryptoassets on different blockchains, and a rather new subset of analytic efforts target cross-currency transactions and their traceability through exchanges such as ShapeShift (Al Jawaheri et al., 2020). The authors in (Harrigan & Fretter, 2016) clustered the addresses of the whole Bitcoin’s blockchain to show the methodology is still suitable despite mixed transactions.

On the other hand, we find works that focus less on the use of networks analytics, and instead apply machine learning-based forensics (we will discuss some methods in detail in section 4).

The authors in (Yin et al., 2019) presented a supervised learning-based approach to de-anonymize the Bitcoin blockchain to predict the type of entities yet not identified. They built classifiers with respect to 12 categories and concluded it is possible to predict the type of an unidentified entity. To do so, they collaborated with the analytic company Chainalysis, the data provider, which had previously clustered, identified, and categorized a considerable number of addresses manually or through clustering techniques. They show two examples, one where they predict on a set of 22 clusters suspected to be related to criminal activities, and another where they classify 153,293 clusters to provide an estimation of Bitcoin activity. On top of this, they concluded it is possible to predict if a cluster belongs to predefined categories such as exchange, gambling, hosted wallet, merchant services, mining pool, mixing, ransomware, scam, tor market.

Machine learning solutions can benefit from constructing multiple graph types from blockchain data, *e.g.*, a blockchain account is a node and a single transaction between two accounts is an edge, or a group of accounts is a node and the edges are defined as the aggregate transaction volume over a period of time. The latter is the predominant forensic method for cryptocurrency activity seen in Section 2.2 (Weber et al., 2018). Relatedly, the authors in (Weber et al., 2019) benchmarked GCN against various supervised methods, while the authors in (Eddin et al., 2021) extended their work with the aim to reduce false alerts through supervised machine learning methods in a non-IoM context.

They call the machine learning component the “triage model”, tasked to process the rule-generated alerts: the score produced by the system enables alert suppression or alert prioritization. The GuiltyWalker (Oliveira et al., 2021) leverages random walks on a cryptocurrency graph to characterize distances to previous suspicious activity. Other works propose graph-based suspiciousness scores based on a detection system incorporating business knowledge about money flows, without the use of learning algorithms (X. Li et al., 2020; Sun et al., 2021).

4 Anomaly Detection Approaches

The process of “anomaly” or “outlier” detection involves the processing of data to detect behaviour patterns that may indicate a change in system operations, thus singling out rare or suspicious (*i.e.*, significantly different from the dataset) events/items (Kamišalić, Kramberger, & Fister, 2021). In this respect, “collective” anomaly detection methods target groups of data points that differ from the majority of the data, while “point” anomaly detection techniques consider also single data points (Z. Li, Xiang, Gong, & Wang, 2022; Shayegan et al., 2022). AML regulated entities, especially in the banking industry, deploy RegTech solutions to screen their operations and detect anomalous activities in an automated way, thus grounding their transaction monitoring procedures on these tools. Red flag indicators are usually provided by regulatory frameworks in a rules-based format, *i.e.*, templates of sequences of actions that suggest a suspicion, in a way that is self-explainable, as required for auditing purposes. Hence, the preliminary review of a flagged account relies on “suspiciousness heuristics” (*e.g.*, political exposure, geographic location, transaction type, users’ behaviour) (Weber et al., 2018), such as FATF’s anomaly indicators related to virtual assets (VAs), developed from analyzing 100+ case studies from 2017 to 2020 (FATF, 2020).

Rules-based red flags can pertain to transaction patterns, such as “*incoming transactions from many unrelated wallets in relatively small amounts (accumulation of funds) with subsequent transfer to another wallet or full exchange for fiat currency*”, or to anonymity, such as “*moving a VA that operates on a public, transparent blockchain, such as Bitcoin, to a centralised exchange and then immediately trading it for an AEC or privacy coin*” (FATF, 2020). In this context, a lot of time and resources are needed to investigate alerts generated by rule-matching processes, to decide whether to submit a suspicious transaction reports (see sub-section 2.2). Because an alert can be a true or a false positive, arguably rules-based systems have the advantage of interpretability, but their simplicity produces too many false positives, estimated at around 95–98% (Eddin et al., 2021).

Indeed, classifying entities and discovering patterns in massive time-series transaction datasets that are dynamic, high dimensional, combinatorially complex, non-linear, often fragmented, inaccurate, incomplete or inconsistent is

not a trivial task. Moreover, the difficulty of automating the synthesis of information from multi-modal data streams thrusts the task onto human analysts. In this case, there is a vicious circle of a compliance approach that stimulates over-reporting because of the cost asymmetry between false positives and false negatives, and overburdens law enforcement agencies (Weber et al., 2018). Hence, the automation of an increasing array of processes has been suggested (Oad et al., 2021).

Against this backdrop, this section lays out the theoretical background on anomaly detection methods applied in our research and elaborates on the methodologies that relate to our experiments. Hence, we focus on machine learning and graph analysis, taking an on-chain data analytic perspective. Nonetheless, we acknowledge the significant role played in the AML context by other tools that target off-chain data, such as Natural Language Processing, sentiment analysis, that also leverage graph methods (Weber et al., 2018).

4.1 Machine Learning

Machine learning is a part of artificial intelligence that exploits data and algorithms to imitate human learning processes, with gradual accuracy improvements. This helps us find solutions to problems in many fields, *e.g.*, vision, speech recognition, robotics (Alpaydin, 2020). In the most diverse contexts, it provides tools that can learn and improve automatically leveraging the vast amount of data available in our age (Kamışalić et al., 2021). In the compliance domain, advances in these algorithms show great promise and their deployment in AML RegTech solutions can improve the efficiency of these applications (Weber et al., 2019). For instance, they can mitigate the shortcomings of rules-based systems and infer patterns from historical data, thus increasing detection rates and decreasing the number of false positives (Lorenz, 2021). In other cases, a more proactive approach is deployed to map and predict illicit transactions (Koshy, Koshy, & McDaniel, 2014; Weber et al., 2019). One of the main distinctions in machine learning is between “unsupervised” methods, where the model works on its own to discover patterns and information previously undetected, and “supervised” techniques, where labeled datasets are used to train algorithms. While it is possible to apply both methods for anomaly detection, most systems deploy unsupervised techniques due to a lack of relevant real-world labeled datasets.

In the AML sphere, this scarcity chiefly derives from difficulties in labeling real cases timely and comprehensively. Indeed, manual labels are costly effort- and time-wise, and the nature of the entities involved is complex and ever-evolving (Lorenz, 2021). In this context, the business line of analytic companies play a crucial role in labelling cryptocurrency transaction data. To address the overall lack of data, various strategies have been proposed (Eddin et al., 2021): generate a fully synthetic dataset, simulate only unusual accounts within a real-world dataset, localize rare events within a peer group. However, better validations of the systems were obtained using analyst feedback or real labeled

data. Parallely, the shortage of real-life datasets has driven the deployment of unsupervised and active learning (*i.e.*, few labels) (Lorenz, 2021).

4.1.1 Supervised Baseline Techniques

Supervised Learning techniques are leveraged for their labeled training data. Some instances of their use include the classification of anomalies based on association rules to detect suspicious events (Luo, 2014), and, in the context of AML, the label of each transaction could indicate whether it was identified as money laundering or not (Lorenz, 2021). Recent RegTech tools make use of widespread Supervised Learning methods to perform anomaly detection (Yin et al., 2019):

- **Decision Tree** - It is one of the base algorithms being used in machine learning, with a name derived from a hierarchical model formed visually as a tree, in which nodes are decisions with a certain criteria. Following the different tree's branches, the training data is then subdivided in different subsets. The decision criteria in the nodes are determined variables that can be defined explanatory and the learning algorithm tries to apply the most significant feature to perform the best division among the training data. The best division can be measured by the information gain which is mathematically derived from a decrease in entropy (Alpaydin, 2020).
- **Random Forests** - Consists of an extension of decision trees in which an algorithm approaches the classification task by constructing a multitude of trees. Introduced by Breiman (Breiman, 2001), it is an ensemble method that is applied to sample random subsets of the training data for each Decision Tree. It aims to improve predictive accuracy of a classifier by combining multiple individual weak learners, *i.e.* trees.
- **Boosting Algorithms** - They are another type of ensemble methods that fit sequences of weak learners. A boosting algorithm tries to boost a Decision Tree by recursively selecting a subset of the training data. AdaBoost (Adaptive Boosting) assigns weights to the samples of data, based on the ability of the weak learners, in order to predict the individual training sample. Thus, for each iteration, the sample weights are individually computed and the successive learner is applied to the new data subset (Yin et al., 2019).
- **Logistic Regression** - It is a multiple regression suitable for binary classification, which assesses the relationship between the binary dependent variable (target) and a set of independent categorical or continuous variables (predictors) (Hilbe, 2009). In principle, the logistic regression can be seen as the measurement of the probability of an event happening. This probability consists of the ratio between the probability that an event will occur and the probability that the event will not occur.
- **Support Vector Classification (SVC)** - Given a set of data for training, each labelled with the class to which it belongs among the two possible classes, a training algorithm for Support Vector Machines builds a model

that assigns the new data to one of the two classes, thus obtaining a non-probabilistic binary linear classifier. This model uses a representation of data as points in space, mapped in such a way that data belonging to the two different categories are clearly separated by as large a space as possible. New data are then mapped in the same space and the prediction of the category to which they belong is made on the basis of the side in which they fall (Alpaydin, 2020).

- **k-Nearest Neighbours (k-NN)** - It is a supervised learning algorithm used in pattern recognition for the classification of objects based on the characteristics of the objects close to the considered one. Also in this case the model uses a representation of data as points in space, i.e. the feature space. Given a notion of distance between data objects, the input is the k nearest training data in the feature space. The underlying idea is that the more similar the instances, the more likely it is that they belong to the same class (Alpaydin, 2020).

4.1.2 Graph Analysis

In recent years, for some areas of machine learning there has been a focus on real-world datasets that come in the form of graphs or networks, e.g., social networks, knowledge graphs, etc., in order to generalize of learning models to such structured datasets. For AML graph analytics is emerging as an increasingly important tool because money laundering involves transactions flow relationships between entities, thus creating a graph structure. Some approaches have been provided for supervised learning on graph-structured data that is based on a variant of neural networks which operate directly on graphs, i.e. Graph Neural Network (Jiaxuan You, 2020; Kipf & Welling, 2016). Convolutional neural networks, for instance, offer an efficient architecture to extract highly meaningful statistical patterns in large-scale and high-dimensional datasets and can be generalized to graphs (Defferrard, Bresson, & Vandergheynst, 2016; Kipf & Welling, 2016).

- **Graph Convolutional Networks (GCN)** - The objective of a GCN model is to learn a function of signals/features on a data set structured as a graph. The model takes as input (i) a graph with nodes and edges between nodes, and (ii) a feature description for each node. The key idea is that each node receives and aggregates features from its neighbours in order to represent and compute its local state. The GCN then usually produces an output feature matrix at the node level (Kipf & Welling, 2016).
- **Graph Attention Networks (GAT)** - While the GCN model averages the node states from source nodes to the target node, the GAT model instead firstly applies normalized attention scores to each source node state and then operates on the basis of these (Veličković et al., 2017). Put in other words, according to a GAT model, given a node different importance to each edge is given, through the introduction of attention coefficients.

5 Experimenting With Machine Learning

We already mentioned that AML analytics could benefit from analysis techniques that use machine learning to classify transactions. In view of this claim, this section outlines the experimental setup of the study we performed and the results obtained. After describing the dataset used in the experiments, the evaluation method and the implementations of the anomaly detection methods are considered. Subsequently, we compare results of our experiments, where state of the art machine learning techniques and graph based neural networks are employed in this AML context.

5.1 Methodology

The starting point for our experimentation was a seminal work available in the state of the art (Weber et al., 2019). Starting from that work, in this section we report on a general classification comparison with a set of benchmark classification methods, *i.e.*, Decision Trees, Logistic Regression, k-NN, SVC, AdaBoost, Random Forests, Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). These two types of neural networks take into account the graph nature of our dataset. As the evaluation will show, the adoption of a novel implementation of GCN improves the performance of such classification scheme.

The dataset is the publicly available Elliptic transactions dataset (Weber et al., 2019), containing real Bitcoin transactions represented as a directed graph network. In such a graph, transactions are nodes, whereas the directed edges between these transactions represent payments flow from the source address to the destination address.

The dataset contains a total of 203769 transactions nodes connected by 234355 edges. For each transaction, 167 features are available, of which the first 94 are relative to the transaction itself and thus extracted from the blockchain directly. An example would be for instance the number of inputs of a transaction or the number of outputs. The following 73 features are relative to the graph network itself and are extracted from the neighbouring transactions of a node. Using GCN it is possible to get a more detailed picture of what information the model has learnt about the nodes and their neighbourhoods, *i.e.*, an embedding of the node into a latent vector space that captures that information and that comes in the form of a look-up table mapping node to a vector of numbers. Tests were carried out both with the transaction features (tx), as well as with transaction features plus aggregated features (tx + agg). Such aggregated features are obtained by aggregating transaction information one-hop backward/forward from the center transaction node. This means obtaining the features of the nodes that share an edge with that transaction node.

Each transaction in the dataset is then labeled as illicit, licit or unknown: 4,545 are labeled as illicit, 42,019 are labeled as licit and the remaining 157205 are unknown. The transactions also contain temporal information. In particular these are grouped in 49 distinct time steps, evenly spaced in the interval

of two weeks. Each time step contains a connected graph containing all the transactions verified on the blockchain in the span of 3 hours (Weber et al., 2019).

We performed a preprocessing on the dataset that can be summarized as follows:

- merge the features with the classes;
- rename class values to integer values;
- swap transaction identifiers for a sorted index;
- select only the part of the dataset labeled licit or illicit;
- remove all the edges between unknown transactions.

After the preprocessing, our cleaned dataset was composed of 46564 transactions and 36624 edges.

5.2 Transaction Graph Analysis

Transactions are linked by nature, since money spent in a transaction originates from previous transfers. This allows the creation of a graph of transactions that can be of help in the classification process. In fact, given a transaction t , it is possible to collect all the connected transactions, and recursively search for other ones up to a certain depth level. Given such connected graph centered at t , inspection of the neighbouring transactions and their classified value can aid the classification of t .

An example of this procedure is reported in Figure 1, where a connected component, *i.e.*, a subgraph in which each pair of nodes is connected with each other via a path, is obtained from an initial transaction (Figure 1, top). In the Figure the red nodes represent transaction labeled as illicit in the starting dataset, green ones the licit transaction and the grey ones the unknowns. In order to show the output of a machine learning classification problem, the bottom part of Figure 1 shows the output of the process employing a specific classification algorithm, *i.e.*, in this case Random Forest. In essence, the idea is that knowing the labels of certain transactions aids the classification of the remaining unknown ones.

The majority of the exploited machine techniques do not deserve a detailed explanation, since their use was standard and their specifics are covered by a wide literature. Among them, however, two specific models deserve close attention in this context, for different reasons. In particular, on the one hand we provide a description of the specific implementation of our GCN model since, as detailed in the results, our implementation allows to obtain results that outperform other approaches and improve the state of the art. On the other hand, we describe also the GAT model since the experiment described in this work is, to the best of our knowledge, the first attempt to use this technique in the AML context.

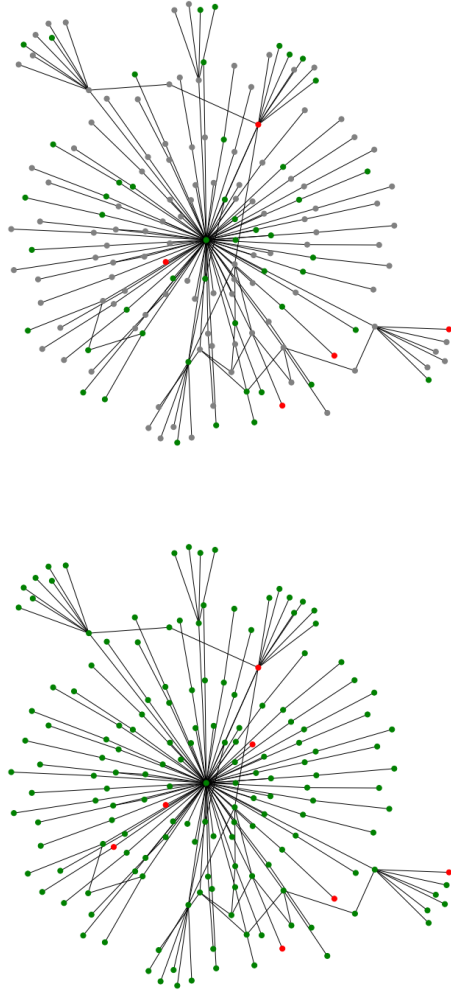


Fig. 1 Connected graph of a considered transaction before and after classification

Table 1 GCN model architecture. Total parameters = 18,774, Trainable parameters = 17,756, Non-Trainable = 1,018.

Layer (type)	Output Shape	Num. parameters
preprocess (Sequential)	(46564, 32)	4564
convolution 1 (GraphConvLayer)	multiple	5888
convolution 2 (GraphConvLayer)	multiple	5888
postprocess (Sequential)	(46564, 32)	2368
logits (Dense)	multiple	66

5.2.1 Graph Convolutional Network Model Architecture

GCN have been developed using the Keras framework, following the recommendations introduced in (Jiaxuan You, 2020). The general structure of our graph convolution layer is made of three steps. First, the input node representations are processed using a Feed Forward Network (FFN) to produce a message. Second, the messages of the neighbours of each node are aggregated using a permutation invariant pooling unsorted segment sum operation. Third, the node representations and aggregated messages are combined and processed to produce the new state of the node representations (node embeddings), via concatenation and FFN processing.

Our network architecture consists of a sequential workflow of the model is summarized as follows (see Table 1):

1. Apply pre-processing using FFN to the node features to generate initial node representations;
2. Apply two graph convolutional layers, with skip connections, to the node representation to produce node embeddings;
3. Apply post-processing using FFN to the node embeddings to generate the final node embeddings;
4. Feed the node embeddings in a Softmax layer to predict the node class.

Table 2 GAT model architecture. Total parameters = 59,952, Trainable parameters = 59,952, Non-Trainable = 0.

Layer (type)	Output Shape	Num. parameters
dense 9 (Dense)	multiple	10340
dropout 6 (Dropout)	multiple	0
graph attention (MultiHeadGraphAttention)	multiple	12320
dense 10 (Dense)	multiple	36630
dropout 7 (Dropout)	multiple	0
dense 11 (Dense)	multiple	662

5.2.2 Graph Attention Network Model Architecture

In GAT model, we use an attention mechanism to aggregate information from neighboring nodes. In other words, instead of simply averaging/summing node states from source nodes to the target node, as we do in the GCN model, GAT instead firstly applies normalized attention scores to each source node state and then sums (Veličković et al., 2017).

Our model is built using the Keras framework, through a graph attention layer that computes pairwise attention scores, aggregates and applies the scores to the node's neighbours. A multi head attention layer concatenates multiple graph attention layer outputs. Our design choice is to use a single attention layer with multiple heads, making the network able to jointly attend multiple positions (Liyuan Liu, 2021). The multi-head layer is then inserted into a general model that implements dense pre-processing/post-processing

layers with dropout regularization, as shown in Table 2. The training proved to be subjected to overfitting and heavy regularization was necessary, which was achieved by dropout layers and with the use of RMSprop optimizer with momentum (Philipp, Song, & Carbonell, 2017).

Table 3 Table showing the results for the F1-score, Micro Average F1-score, Precision and Recall metrics for all models

Model	Precision	Recall	F1 Score	M.A. F1
Random Forest Classifier (tx)	0.909	0.648	0.757	0.974
Random Forest Classifier (tx + agg)	0.981	0.651	0.782	0.977
Logistic Regression (tx)	0.515	0.646	0.573	0.939
Logistic Regression (tx + agg)	0.456	0.630	0.529	0.929
MLP (tx)	0.897	0.593	0.714	0.970
MLP (tx + agg)	0.817	0.623	0.707	0.968
k-NN Classifier (tx)	0.762	0.629	0.689	0.964
k-NN Classifier (tx + agg)	0.730	0.576	0.644	0.960
SVC (tx)	0.842	0.604	0.703	0.968
SVC (tx + agg)	0.862	0.588	0.699	0.968
Decision Tree Classifier (tx)	0.986	0.573	0.725	0.973
Decision Tree Classifier (tx + agg)	0.986	0.573	0.725	0.973
AdaBoost Classifier (tx)	0.793	0.615	0.693	0.966
AdaBoost Classifier (tx + agg)	0.945	0.567	0.708	0.971
GCN (tx)	0.906	0.790	0.844	0.973
GAT (tx)	0.897	0.605	0.723	0.971

5.3 Results

While we considered both licit and illicit classes, for the sake of readability and conciseness in this Section we focus specifically on F1-score for the illicit class. F1-score is a metrics obtained from Precision and Recall. Precision is the number of true positives over the amount of identified (true and false) positives and the Recall is the number of true positives over the sum of true positives and false positives. The F1-score represents the harmonic mean of Recall and Precision and is thus calculated as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

We also use the Micro Average F1-score for evaluating the methods. It measures the F1-score of the aggregated contributions of all classes.

The final performance results are reported in Figure 2. It is possible to observe how GCN outperforms other approaches. In particular, the GCN approach provides the best results in terms of recall, *i.e.*, 0.790, and F1-score, *i.e.*, 0.844. In terms of precision, it slightly deviates from the Decision Tree (0.986) and Random Forest (0.981) approaches, but still provides better performances than all the rests, *i.e.*, 0.906. For what concerns the Micro Average F1-score, all approaches fit in the range of 0.960 to 0.977. These results are in

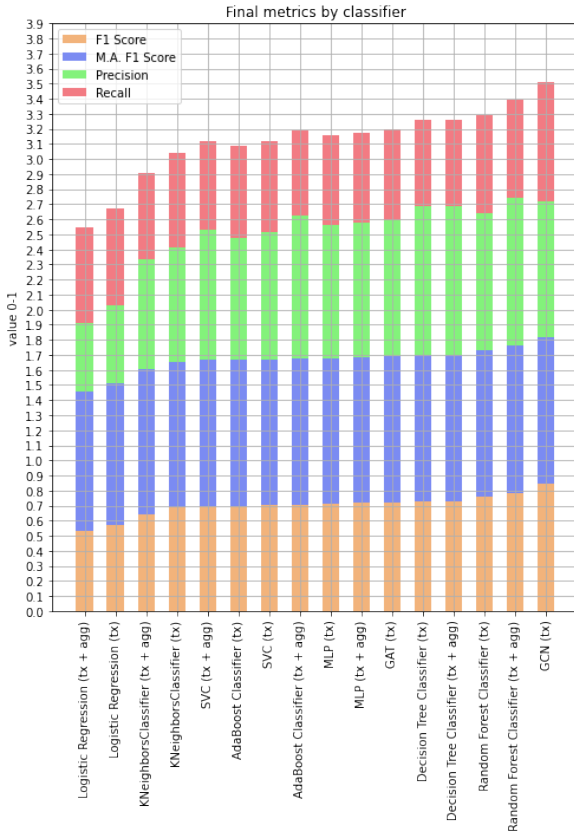


Fig. 2 Barplot aggregating F1-score, Micro Average F1-score, Precision and Recall for all the approaches experimented.

contrast with the results in (Weber et al., 2019), where Random Forests provided best performance. The motivation behind such improvement might be due to the different architecture we exploited to build the neural network.

Furthermore, the comparison with the other graph-based approach, *i.e.*, GAT, also sees the GCN outperforming. GAT performs better than simple dense network but it is not able to reach the results of GCN and Random Forest classifiers. The motivation is probably due to a naive structure of the neural network and its optimization is currently under investigation.

6 Discussion

The experiment described in this paper is, to the best of our knowledge, the first attempt at implementing GAT models to detect anomalies on Bitcoin for AML purposes. The final results are on par with the state of the art of GCN networks, with GAT being marginally worse than GCN. This could be explained by the “simpler” implementation of GAT, the possibility that the

dataset responds better to non-spectral methods. Nonetheless, we argue that the novelty of this application could be useful for general research on GAT anomaly detection techniques. In addition, the results show that the GCN neural network typology is also a promising solution for AML, as it performs better than other approaches.

One has to consider that GCN and GAT classifiers only have access to transaction features, which means that all information about aggregated nodes comes from the graph structure itself. Considering that the performance of GCN is in line with Random Forest (with aggregated features) we can claim that our graph networks are able to obtain the same amount of information as the creator of the dataset (Weber et al., 2019).

The choice of one method over another, however, carries additional implications to consider. For example, Random Forests' performance falls slightly behind GCN's results, yet because the detectors are derived from Random Forests' rules, there is no sacrifice in explainability. (Eddin et al., 2021). Given the size and dynamism of real-world information, results' explainability is difficult to provide, both in this context and in the broader field of artificial intelligence. Even in our case, which is a narrow instance, *i.e.*, with transaction graphs that model illicit activity over time, it is difficult to apply methods that are efficient and whose results can be understood by humans. Although this appears to be a crucial aspect, the literature is still lacking some research in the area of the application of explainable artificial intelligence techniques to the domain of AML detection (Kute, Pradhan, Shukla, & Alamri, 2021).

Meanwhile, the deployed forensic approaches must be considered against the backdrop of a substantial evolution of the IoM, which warrants for the application of increasingly sophisticated, yet explainable, compliance and investigation techniques. In this respect, as reported above the "crypto travel rule" has stimulated the industry to denounce the lack of global standards and technical solutions to underpin effective and affordable AML compliance. As mentioned in the previous sections, the great quantity and complexity of transaction data to be processed, today and even more in the future, suggests that machine learning techniques will continue to be a paramount part of the solution. Hence, even marginal performance differences may bear a significant weight, especially when considered under the lens of possible combinations between different approaches.

Meanwhile, the lack of readily available "travel rule" compliance solutions goes beyond the specific "travel rule" context, and displays how regulation is still largely targeting a centralized exchange (CEX) model and has yet to conceive an approach that suits peer-to-peer transfers and decentralized exchanges (DEXes). Even if one assumes DEXes (or a subset of them) fall within the scope of AML frameworks – which is *per se* debatable –, after launch they are usually operated automatically by the protocol and governance tokens holders. Hence, it is a problem to identify an entity on which to impose compliance, while a CEX-centric AML approach keeps shifting liquidity to DEXes. Indeed, rising DEX misuse was detected in terms of an increase in laundering-related DeFi

protocols' usage of 1.964% between 2020 and 2021. In 2021, CEXes received 47% of funds originating from illicit addresses and DeFi protocols 17%, vis-à-vis 2% in 2020. Likewise, in 2021 funds derived from cryptocurrency thefts were increasingly sent to DeFi platforms (51%) or risky services (25%), while only 15% went to CEXes, possibly due to AML ([Chainalysis Team, 2022](#)).

Concurrently, standardization loopholes arise on a regular basis due to emerging IoM-related obfuscation techniques or increase in their use (*e.g.*, chain-hopping and dusting, atomic swapping exchanges, privacy wallet-based transfers combining multiple transactions into a single one, such as Coin-Join) ([FATF, 2021a, 2021b](#)). This type of environmental developments are far from being unprecedented. Indeed, the first AML focus on cryptocurrencies considered only “on and off ramps” to the traditional (regulated) financial system (*i.e.*, fiat-to-crypto exchanges). Later, however, the international focus broadened to include virtual-to-virtual schemes ([FATF, 2019](#)).

Given the evolution towards enhanced disintermediation, regulated entities and law enforcement agencies will have an increasing need of analyzing a large amount of transactions not only subject to sophisticated obfuscation techniques, but also without the assistance of regulated entities as counterparties – *e.g.*, investigations or customer due diligence (see sub-section 2.2) on transactions originating from or destined to self-hosted wallets or processed by DEXes. For this reason, beside highlighting the value of researching innovative applications of forensic methods based on machine learning, we argue the adoption of an interdisciplinary approach is crucial to devise compliance tools that take into account the way regulatory regimes are conceived and enforced. In this regard, our work performs a contextualization of forensic methods into the specifics of the rules-based anomaly indicators provided by the AML framework (*e.g.*, FATF's VA-related red flags). In this context, we consider the role of machine learning in mitigating the shortcomings of rules-based systems, while advocating for the development of cross-disciplinary models to aid compliance and reduce over-reporting.

7 Conclusions

While blockchain technology became popular in the wake of Bitcoin's launch, experts and law enforcement professionals devised intelligence strategies to trace cryptocurrency transfers. Blockchain forensics techniques started to be deployed to connect blockchain addresses and transactions to real-world identities in order to combat money laundering and the financing of terrorism and proliferation (AML/CFT). Cryptocurrency misuse on the darknet prompted an extension of the scope of application of the AML framework, internationally overseen by the FATF, to include IoM's transactions. Likewise, relevant anomaly indicators were drafted to help regulated entities to identify suspicious transfers. In this context, investigating authorities and law enforcement agencies began to successfully apply forensic methods to track currency

across blockchain ecosystems, while regulated entities were offered innovative RegTech solutions that detect anomalous activities in an automated way.

In this paper, our focus has been on the growing relevance of such techniques and in particular on machine learning and graph analysis approaches, taking an on-chain data analytic perspective. In the compliance domain, advances in these algorithms show great promise and their use in RegTech AML solutions can improve the efficiency of these applications. We have performed an experimentation that, as far as we know, is the first one involving the use of Graph Attention Networks (GAT) models on AML anomaly detection in Bitcoin. This type of neural network stems from a recent focus on exploiting the fact that many real-world datasets come in the form of graphs or networks. Thus, Graph Convolutional Networks (GCN) and GAT were born with the idea of creating generalized learning models for such structured datasets. Indeed, the data we analyzed consist of a real-world dataset of Bitcoin transactions represented as a directed graph network.

The final results show that these graph-based approaches perform better than the baseline approaches, *e.g.*, GCN performs better than Random Forests, with GAT being marginally worse than GCN. This suggests that the use of GCN neural networks is also a promising solution for AML, and that the novelty of our approach could be useful for further research into GAT anomaly detection techniques. We therefore conclude that experiments in different applications of the widest range of forensic tools, possibly leveraging the added value of transaction graphs, are crucial to boost relevant research and engage different community stakeholders in a constructive dialogue.

Acknowledgements

The contribution of Nadia Pocher and Mirko Zichichi received funding from the EU H2020 research and innovation programme under the MSCA ITN European Joint Doctorate grant agreement No 814177 Law Science and Technology Joint Doctorate - Rights of the Internet of Everything.

References

- Akhgar, B., Gercke, M., Vrochidis, S., Gibson, H. (2021). *Dark Web Investigation*. Springer. 10.1080/09546553.2021.2000216
- Al Jawaheri, H., Al Sabah, M., Boshmaf, Y., Erbad, A. (2020). Deanonymizing Tor hidden service users through Bitcoin transactions analysis. *Computers and Security*, 89. <https://arxiv.org/abs/1801.07501> 10.1016/j.cose.2019.101684
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Amarasinghe, N., Boyen, X., McKague, M. (2019). A Survey of Anonymity of Cryptocurrencies. *Acm international conference proceeding series*. Sydney: Association for Computing Machinery. 10.1145/3290688.3290693

Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T., Capkun, S. (2013). Evaluating user privacy in Bitcoin. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7859, 34–51.

10.1007/978-3-642-39884-1_4

Antonopoulos, A.M. (2017). *The internet of money - volume two*. Merkle Boom LLC.

Authors blinded for review (n.d.). *Published work*.

Bartoletti, M., Carta, S., Cimoli, T., Saia, R. (2020). Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact. *Future Generation Computer Systems*, 102, 259–277. <https://arxiv.org/abs/1703.03779>
10.1016/j.future.2019.08.014

Berg, A. (2019). The Identity, Fungibility and Anonymity of Money. *Economic Papers*(November), 1–16. Retrieved from <https://ssrn.com/abstract=3211011>

10.1111/1759-3441.12273

Biryukov, A., & Tikhomirov, S. (2019). Deanonymization and linkability of cryptocurrency transactions based on network analysis. *Proceedings - 4th IEEE European Symposium on Security and Privacy, EURO S and P 2019*, 172–184.

10.1109/EuroSP.2019.00022

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.

Chainalysis Team (2022). The 2022 Crypto Crime Report. (February).

Chen, W., Zheng, Z., Ngai, E.C., Zheng, P., Zhou, Y. (2019). Exploiting Blockchain Data to Detect Smart Ponzi Schemes on Ethereum. *IEEE Access*, 7(c), 37575–37586. Retrieved from https://www.researchgate.net/publication/331853833_Exploiting_Blockchain_Data_to_Detect_Smart_Ponzi_Schemes_on_Ethereum

10.1109/ACCESS.2019.2905769

- CipherTrace (2021). *Cryptocurrency Crime and Anti-Money Laundering Report* (Tech. Rep. No. February).
- Custers, B., & Overwater, L. (2019). Regulating Initial Coin Offerings and Cryptocurrencies: A Comparison of Different Approaches in Nine Jurisdictions Worldwide. *European Journal of Law and Technology*, 10(3).
- Defferrard, M., Bresson, X., Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29.
- Directive (EU) 2018/843 (2018). *Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU*.
- Eddin, A.N., Bono, J., Aparício, D., Polido, D., Ascensão, J.T., Bizarro, P., Ribeiro, P. (2021). *Anti-money laundering alert optimization using machine learning with graphs*. arXiv. Retrieved from <https://arxiv.org/abs/2112.07508> 10.48550/ARXIV.2112.07508
- Edmunds, J.C. (2020). *Rogue Money and the Underground Economy. An Encyclopedia of Alternative and Cryptocurrencies*. Greenwood, ABC-CLIO.
- European Commission (2021). *Anti-money laundering and countering the financing of terrorism legislative package*. Retrieved from https://ec.europa.eu/info/publications/210720-anti-money-laundering-counter-terror-finance_en
- Europol (2020, oct). *Internet Organised Crime Threat Assessment 2020* (Tech. Rep.). Retrieved from <https://www.europol.europa.eu/>
- FATF (2019). *Guidance for a risk-based approach: virtual assets and virtual asset service providers* (Tech. Rep. No. June). Paris: Author. Retrieved from www.fatf-gafi.org/
- FATF (2020). *Virtual Assets Red Flag Indicators of Money Laundering and Terrorist Financing* (Tech. Rep. No. September). Retrieved from <http://www.fatf-gafi.org/>
- FATF (2021a). *Second 12-Month Review of the Revised Fatf Standards on Virtual Assets and Virtual Asset Service Providers*. (July). Retrieved from

<https://www.fatf-gafi.org/media/fatf/documents/recommendations/Second-12-Month-Review-Revised-FATF-Standards-Virtual-Assets-VASPS.pdf>

FATF (2021b). Virtual Assets and Virtual Asset Service Providers - Updated Guidance for a Risk-Based Approach. (October). Retrieved from <https://www.fatf-gafi.org/media/fatf/documents/recommendations/Updated-Guidance-VA-VASP.pdf>

FATF (2022, mar). *International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation: the FATF Recommendations* (Tech. Rep.). Retrieved from <https://www.fatf-gafi.org/media/fatf/documents/recommendations/pdfs/FATFRecommendations2012.pdf>

Filippi, P.D., & Wright, A. (2018). *Blockchain and the law: The rule of code*. Harvard University Press. Retrieved from <http://www.jstor.org/stable/j.ctv2867sp>

Fleder, M., Kester, M.S., Pillai, S. (2015). Bitcoin Transaction Graph Analysis. , 1–8. Retrieved from <http://arxiv.org/abs/1502.01657> <https://arxiv.org/abs/1502.01657>

Goforth, C.R. (2020). Crypto Assets : A Fintech Forecast. (September), 5–25.

Harrigan, M., & Fretter, C. (2016). The unreasonable effectiveness of address clustering. *2016 intl ieee conferences on ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (uic/atc/scalcom/cbdcom/iop/smartworld)* (pp. 368–373). IEEE.

Harvey, J., & Branco-Illodo, I. (2020). Why Cryptocurrencies Want Privacy: A Review of Political Motivations and Branding Expressed in “Privacy Coin” Whitepapers. *Journal of Political Marketing*, 19(1-2), 107–136. Retrieved from <https://doi.org/10.1080/15377857.2019.1652223>

10.1080/15377857.2019.1652223

Hilbe, J.M. (2009). *Logistic regression models*. Chapman and hall/CRC.

Ince, P., Liu, J.K., Zhang, P. (2018). *Adding confidential transactions to cryptocurrency IOTA with bulletproofs* (Vol. 11058 LNCS). Springer International Publishing. Retrieved from <http://dx.doi.org/10.1007/>

978-3-030-02744-5_3 10.1007/978-3-030-02744-5_3

Jiaxuan You, J.L., Rex Ying (2020). Design space for graph neural networks. *arXiv:2011.08843*.

<https://doi.org/10.48550/arXiv.2011.08843>

Johnstone, S. (2021). *Rethinking the Regulation of Cryptoassets. Cryptographic Consensus Technology and the New Prospect*. Elgar.

Kamišalić, A., Kramberger, R., Fister, I. (2021). Synergy of blockchain technology and data mining techniques for anomaly detection. *Applied Sciences (Switzerland)*, 11(17).

10.3390/app11177987

Karasek-wojciechowicz, I. (2021). Reconciliation of anti-money laundering instruments and European data protection requirements in permissionless blockchain spaces. *Journal of Cybersecurity*, 1–28.

10.1093/cybsec/tyab004

Kipf, T.N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Koshy, P., Koshy, D., McDaniel, P. (2014). An Analysis of Anonymity in Bitcoin Using P2P Network Traffic. *International financial cryptography association 2014* (Vol. 8437, pp. 469–485). 10.1007/978-3-662-45472-530

Kute, D.V., Pradhan, B., Shukla, N., Alamri, A. (2021). Deep learning and explainable artificial intelligence techniques applied for detecting money laundering—a critical review. *IEEE Access*.

Li, X., Liu, S., Li, Z., Han, X., Shi, C., Hooi, B., ... Cheng, X. (2020). Flowscope: Spotting money laundering based on graphs. *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 4731–4738).

Li, Y., Susilo, W., Yang, G., Yu, Y., Du, X., Liu, D., Guizani, N. (2019). Toward privacy and regulation in blockchain-based cryptocurrencies. *IEEE Network*, 33(5), 111–117.

10.1109/MNET.2019.1800271

Li, Y., Yang, G., Susilo, W., Yu, Y., Au, M.H., Liu, D. (2021). Traceable Monero: Anonymous Cryptocurrency with Enhanced Accountability. *IEEE Transactions on Dependable and Secure Computing*, 18(2), 679–691.

10.1109/TDSC.2019.2910058

Li, Z., Xiang, Z., Gong, W., Wang, H. (2022). Unified model for collective and point anomaly detection using stacked temporal convolution networks. *Applied Intelligence*, 52(3), 3118–3131. Retrieved from <https://doi.org/10.1007/s10489-021-02559-0>

10.1007/s10489-021-02559-0

Lischke, M., & Fabian, B. (2016). Analyzing the bitcoin network: The First Four Years. *Future Internet*, 8(1).

10.3390/fi8010007

Liyuan Liu, J.H., Jialu Liu (2021). Multi-head or single-head? an empirical comparison for transformer training. *arXiv:2106.09650*.

<https://doi.org/10.48550/arXiv.2106.09650>

Lorenz, J.S. (2021). *Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity* (Unpublished doctoral dissertation).

Luo, X. (2014). Suspicious transaction detection for anti-money laundering. *International Journal of Security and Its Applications*, 8(2), 157–166.

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S. (2016). A fistful of Bitcoins: Characterizing payments among men with no names. *Communications of the ACM*, 59(4), 86–93.

10.1145/2896384

Moreno-Sanchez, P., Zafar, M., Kate, A. (2016, 02). Listening to whispers of ripple: Linking wallets and deanonymizing transactions in the ripple network. *Proceedings on Privacy Enhancing Technologies*, 2016.

10.1515/popets-2016-0049

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. *www.bitcoin.org*.

Neudecker, T., & Hartenstein, H. (2017). Could network information facilitate address clustering in bitcoin? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10323 LNCS, 155–169.

10.1007/978-3-319-70278-0_9

Oad, A., Razaque, A., Tolemysov, A., Alotaibi, M., Alotaibi, B., Zhao, C. (2021). Blockchain-enabled transaction scanning method for money laundering detection. *Electronics*, 10(15). Retrieved from <https://www.mdpi.com/2079-9292/10/15/1766>

10.3390/electronics10151766

Ober, M., Katzenbeisser, S., Hamacher, K. (2013). Structure and anonymity of the bitcoin transaction graph. *Future Internet*, 5(2), 237–250.

10.3390/fi5020237

Oliveira, C., Torres, J., Silva, M.I., Aparício, D., Ascensão, J.T., Bizarro, P. (2021). Guiltywalker: Distance to illicit nodes in the bitcoin network. *arXiv preprint arXiv:2102.05373*.

Pfitzmann, A., & Hansen, M. (2010). A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. *Technical University Dresden*, 1–98.

10.1.1.154.635

Phan, T. (2021). *Exploring Blockchain Forensics*.

Philipp, G., Song, D., Carbonell, J.G. (2017, December). The exploding gradient problem demystified - definition, prevalence, impact, origin, tradeoffs, and solutions. <https://arxiv.org/abs/1712.05577> [cs.LG]

Quiniou, M. (2019). *Blockchain: the advent of disintermediation*. ISTE Ltd.

Reid, F., & Harrigan, M. (2013). An analysis of anonymity in the bitcoin system. *Security and Privacy in Social Networks*, 197–223. <https://arxiv.org/abs/1107.4524>

10.1007/978-1-4614-4139-7_10

Shayegan, M.J., Sabor, H.R., Uddin, M., Chen, C.-L. (2022). A collective anomaly detection technique to detect crypto wallet frauds on bitcoin

network. *Symmetry*, 14(2). Retrieved from <https://www.mdpi.com/2073-8994/14/2/328>

10.3390/sym14020328

Sun, X., Zhang, J., Zhao, Q., Liu, S., Chen, J., Zhuang, R., ... Cheng, X. (2021). Cubeflow: Money laundering detection with coupled tensors. *Pacific-asia conference on knowledge discovery and data mining* (pp. 78–90).

Tapscott, D., & Euchner, J. (2019). Blockchain and the Internet of Value: An Interview with Don Tapscott Don Tapscott talks with Jim Euchner about blockchain, the Internet of value, and the next Internet revolution. *Research Technology Management*, 62(1), 12–19. Retrieved from <https://doi.org/10.1080/08956308.2019.1541711>

10.1080/08956308.2019.1541711

Tennant, L. (2017). Improving the Anonymity of the IOTA Cryptocurrency. , 1–20. Retrieved from <https://pdfs.semanticscholar.org/490d/https://laurecettenant.com/papers/anonymity-iota.pdf.pdf>

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, F., & De Filippi, P. (2020). Self-Sovereign Identity in a Globalized World: Credentials-Based Identity Systems as a Driver for Economic Inclusion. *Frontiers in Blockchain*, 2(January), 1–22.

10.3389/fbloc.2019.00028

Weber, M., Chen, J., Suzumura, T., Pareja, A., Ma, T., Kanezashi, H., ... Schardl, T.B. (2018). Scalable Graph Learning for Anti-Money Laundering: A First Look. (1970). Retrieved from <http://arxiv.org/abs/1812.00076> <https://arxiv.org/abs/1812.00076>

Weber, M., Domeniconi, G., Chen, J., Weidele, D.K.I., Bellei, C., Robinson, T., Leiserson, C.E. (2019). Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. (10). Retrieved from <http://arxiv.org/abs/1908.02591> <https://arxiv.org/abs/1908.02591>

Wu, J., Liu, J., Chen, W., Huang, H., Zheng, Z., Zhang, Y. (2020, jan). Detecting Mixing Services via Mining Bitcoin Transaction Network with Hybrid Motifs. *Journal of Latex Class Files*, 14(8). Retrieved from

<http://arxiv.org/abs/2001.05233> <https://arxiv.org/abs/2001.05233>

Wu, Y., Tao, F., Liu, L., Gu, J., Panneerselvam, J., Zhu, R., Shahzad, M.N. (2021). A bitcoin transaction network analytic method for future blockchain forensic investigation. *IEEE Transactions on Network Science and Engineering*, 8(2), 1230–1241.

10.1109/TNSE.2020.2970113

Xu, J.J. (2016). Are blockchains immune to all malicious attacks? *Financial Innovation*, 2(1), 25. Retrieved from <https://doi.org/10.1186/s40854-016-0046-5>

10.1186/s40854-016-0046-5

Yin, H.H.S., Langenheldt, K., Harlev, M., Mukkamala, R.R., Vatrappu, R. (2019, jan). Regulating Cryptocurrencies: A Supervised Machine Learning Approach to De-Anonymizing the Bitcoin Blockchain. *Journal of Management Information Systems*, 36(1), 37–73.

10.1080/07421222.2018.1550550