Detecting Potential Fraudulent Trading Patterns in Decentraland: A Data-Driven Analysis of the Ethereum Network and OpenSea Marketplace

Hesam Sarkhosh
University of Waterloo
Waterloo, ON, Canada
hsarkhos@uwaterloo.ca

Abstract

Non-Fungible Tokens (NFTs) are unique digital assets secured and authenticated by blockchain technology, making them vulnerable to fraudulent activities due to lack of central oversight. This study analyzes transactions within Decentraland's virtual land market using the IITP-VDLand dataset [4], which includes comprehensive data on parcel characteristics, trading history, and transaction details. We address key research questions regarding peak periods of activity, influential buyers and sellers, the impact of Ethereum gas fees on transaction volumes, and profit patterns from resale activities. Our findings reveal significant patterns and trends, offering valuable information on the economics of virtual worlds and highlighting potential risks in NFT markets. As the second part of this research, we apply the GTN2vec algorithm [14] to the IITP-VDLand dataset. This method aims to detect potential money laundering patterns within the dataset, using graph-based embeddings to identify anomalous and suspicious activities. The combined analysis provides a deeper understanding of transactional behaviors in NFT markets and offers a framework for mitigating financial crimes in virtual economies.

1 Introduction

Non-Fungible Tokens (NFTs) are unique digital assets secured by blockchain technology, enabling ownership and transfer of digital items ranging from art to virtual real estate. Despite their revolutionary potential, the lack of robust oversight and regulation makes NFTs susceptible to fraudulent activities and market manipulation. This paper focuses on Decentraland, a prominent virtual reality platform where users can buy, sell, and trade virtual land using NFTs. By analyzing transactions on Decentraland through the IITP-VDLand dataset [4], this study aims to uncover patterns in trading activities, identify key players, assess the impact of Ethereum gas fees, and explore profit-making strategies in the virtual land market. [13].

In some cases, these are virtual representations of existing real-world assets such real estate. They could also be assets held within a metaverse such as decentraland's. According to [18], the market has experienced enormous and explosive growth in recent years. The estimated total transaction volume on OpenSea, the largest marketplace, reached 34.7 billion in February 2023.

Non-Fungible Tokens (NFTs) have emerged as a revolutionary class of digital assets, transforming how ownership and value are perceived in the digital realm. Secured by cryptographic technologies and authenticated via blockchain, NFTs have facilitated creating and exchanging unique digital items, including virtual real estate. This paper delves into the transactions of Decentraland on Opensea to explore the dynamics of the virtual land market within the Metaverse.

To provide a comprehensive understanding, we first review some background concepts, including Blockchain and Web3, NFTs, the Metaverse on Web3, Opensea, and Decentraland.

1.1 Background

1.1.1 Blockchain and Web3. Blockchain technology forms the backbone of Web3, the decentralized iteration of the internet. Unlike traditional centralized systems, blockchain operates on a distributed ledger system where data is maintained across a network of computers, ensuring transparency, security, and immutability. Web3 leverages blockchain to create decentralized applications (dApps) that operate without intermediaries, empowering users with greater control over their digital interactions and assets.

1.1.2 NFT. Non-Fungible Tokens (NFTs) are unique digital assets verified using blockchain technology. Unlike fungible assets like cryptocurrencies, each NFT has distinct information and value, making them irreplaceable. NFTs can represent a wide range of digital items, from art and music to virtual real estate and in-game assets. They are typically created and managed on blockchain platforms that support smart contracts, such as Ethereum.

1.1.3 Metaverse on Web3. The Metaverse refers to a collective virtual shared space, created by the convergence of virtually enhanced physical reality and physically persistent virtual spaces. Powered by Web3 technologies, the Metaverse offers a decentralized environment where users can create, own, and monetize digital assets. Within the Metaverse, platforms like Decentraland provide users with immersive experiences where they can engage in activities ranging from gaming and socializing to trading virtual real estate.

1.1.4 OpenSea.

OpenSea is the largest decentralized marketplace for buying, selling, and trading NFTs. It supports a wide array of digital assets, including art, domain names, virtual worlds, trading cards, and collectibles. By leveraging blockchain technology, OpenSea provides a transparent and secure platform for NFT transactions, enabling users to engage in peer-to-peer trading with minimal risk of fraud.

1.1.5 Decentraland.

Decentraland is a virtual reality platform powered by the Ethereum blockchain. Users can create, experience, and monetize content and applications [10]. Land in Decentraland is permanently owned by the community, giving them full control over their creations. Users claim ownership of virtual land on a blockchain-based ledger of parcels. Landowners control what content is published on their portion of land, identified by a set of Cartesian coordinates (x,y). Content can range from static 3D scenes to interactive systems such as games.

Land in Decentraland is a non-fungible, transferrable, scarce digital asset stored in an Ethereum smart contract. It can be acquired by spending an ERC20 token called MANA. MANA can also be used to make purchases of digital goods and services in this digital world. Each parcel in Decentraland is a unique digital asset that can be customized with interactive experiences, 3D scenes, and other digital content. Governed by a Decentralized Autonomous Organization (DAO), Decentraland offers users full control over their creations and the development of the virtual world.

1.2 Motivation

The rapid development of blockchain and Web3 technologies has given rise to the decentralized Metaverse, attracting a substantial influx of users and capital. However, the lack of industry standards and regulatory rules within this ecosystem has made it susceptible to a variety of financial crimes. According to Wu et al. (2022), the decentralized nature of the Web3 Metaverse has witnessed various financial crimes such as scams, code exploits, wash trading, money laundering, and illegal services and shops [20]. The anonymity and lack of oversight inherent in these technologies create fertile ground for malicious activities, posing significant risks to both individual users and the broader financial system.

In the Web3 Metaverse, scams such as Ponzi schemes, rug pulls, phishing attacks, fake exchanges, and giveaway scams are prevalent. These fraudulent activities not only result in financial losses but also undermine trust in the system [20]. Code exploits, where vulnerabilities in blockchain protocols or smart contracts are exploited, pose another significant threat, as evidenced by several high-profile attacks resulting in substantial financial losses [20].

Wash trading, a form of market manipulation where assets are repeatedly traded to create misleading information about market activity is another common crime in the Metaverse. This practice inflates trading volumes and can mislead investors [20].

Money laundering, where illicit funds are channeled through complex financial transactions to disguise their origin, is also facilitated by the anonymous and decentralized nature of blockchain technology. [20].

Given these substantial risks, it is imperative to understand the patterns and mechanisms of financial crimes within the Web3 Metaverse.

In this research, we focus specifically on money laundering, one of the most impactful crimes in the DeFi ecosystem due to its largely unregulated nature. An effective approach to preventing money laundering schemes is by utilizing data to recognize patterns and detect anomalies. Our study aims to explore these approaches by analyzing transactions in Decentraland on OpenSea, leveraging the comprehensive IITP-VDLand dataset. This research seeks to identify vulnerabilities within the system and provide insights into the economic dynamics of the virtual land market, thereby contributing to the development of more robust security measures and regulatory frameworks to safeguard the integrity of the Metaverse.

As a foundational contribution, this paper undertakes a comprehensive analysis of the dataset, identifying and examining key patterns inherent in the data. Through systematic exploration and visualization, we aim to derive valuable insights and provide a detailed understanding of the underlying structures and trends within the dataset.

Building on this, the second major contribution of this paper is the application of the GTN2vec [14] model to the dataset, aiming to detect potential money laundering patterns. This demonstrates the effectiveness of data-driven approaches in identifying suspicious activities, offering a practical framework for tackling money laundering within the Web3 Metaverse.

1.3 Money laundering in Web3

While traditional financial systems offer various means and techniques for money laundering, the Web3 ecosystem—commonly referred to as Decentralized Finance (DeFi) or Finance 2.0—has shown to have its own unique set of methods and tools exploited for facilitating illicit activities. In this article, we explore some of the techniques and mechanisms used for money laundering within the Web3 ecosystem, in simple terms.

1.3.1 Cryptocurrency Tumblers (Mixers).

- Cryptocurrency tumblers or mixers are services designed to improve the anonymity of cryptocurrency transactions. They work by pooling funds from many users and then sending random amounts to different addresses, obscuring the original source of the funds.

- **How it works**:
  1. Source A sends cryptocurrency (e.g., Bitcoin, Ether) to a mixer.
  2. The mixer combines funds from multiple sources and sends them back to different addresses.
  3. The destination (Source B) receives the funds, but the link between A and B is hidden due to the mixing process.
- **Examples**:
  - **Tornado Cash**: One of the most popular decentralized, non-custodial privacy solutions for Ethereum and ERC-20 tokens. It uses zero-knowledge proofs (zk-SNARKs) to ensure users' privacy by mixing their funds.
  - **ChipMixer:** A Bitcoin tumbler that works by providing users with "chips" of different denominations, which are then mixed with other users' chips. This mixing obscures the origin of the funds.
- **Risk**: Many tumblers have been flagged or banned by authorities, and using them might attract scrutiny.

1.3.2 Cryptocurrency with Built-in Privacy (Privacy Coins).

- Some cryptocurrencies have built-in privacy features, making it nearly impossible to trace transactions.
- **Examples**:
  - **Monero (XMR)**: Uses technologies like Ring Signatures, Stealth Addresses, and Confidential Transactions to hide the sender, receiver, and transaction amount.
  - **Zcash (ZEC)**: Uses zk-SNARKs (zero-knowledge proofs) to encrypt transaction details.
- **How it works**:
  1. Source A exchanges their funds (e.g., Bitcoin) for a privacy coin like Monero.
  2. They send Monero to Source B, and Source B exchanges it back to another cryptocurrency or keeps it in Monero.

1.3.3 Layering through Multiple Exchanges.

- This method involves moving funds across multiple exchanges to create a complex trail.
- **How it works**:
  1. Source A sends cryptocurrency to one exchange (Exchange 1).
  2. They trade or convert the funds (e.g., from Bitcoin to Ether), then withdraw it to another exchange (Exchange 2).
  3. This process is repeated across several exchanges, sometimes using different types of cryptocurrencies.
  4. Finally, the funds are sent to Source B, breaking the chain of ownership tracking.
- **Risk**: Many exchanges enforce KYC (Know Your Customer) and AML (Anti-Money Laundering) regulations, which may expose identities.

1.3.4 Using DeFi.

- DeFi protocols allow users to move funds without intermediaries, making tracking harder.
- **How it works**:
  1. Source A deposits funds into a DeFi protocol (e.g., a liquidity pool, yield farming, or staking platform).
  2. The funds are used within the protocol, making it harder to trace their flow.
  3. After some time, the funds are withdrawn to a different address (Source B), making it harder to connect the original source.
- **Risk**: DeFi protocols are more difficult to regulate, but some can still be audited and traced if blockchain forensic techniques are advanced enough.

1.3.5 Offshore Shell Companies, Exchanges, and Banking.

- Using traditional banking, criminals might move funds between offshore bank accounts and companies. They may also use exchanges located in regions where regulators have no access or control.
- **Risk**: This method often relies on exploiting gaps in international financial regulation but can be exposed if authorities request cooperation from multiple jurisdictions.

1.3.6 Smurfing.

- This method involves breaking large amounts of money into smaller, less noticeable transactions.
- **How it works**:
  1. Source A breaks down a large sum into smaller amounts and sends them to different accounts, often across different platforms or currencies.
  2. Source B later consolidates these small amounts.
- **Risk**: Authorities often track smurfing patterns and investigate linked transactions if discovered.

1.3.7 Smart Contracts.

- Smart contracts can be designed to add layers of obfuscation to transactions.
- **How it works**:
  1. Source A transfers funds to a smart contract that automatically moves funds between different addresses or executes trades across different protocols.
  2. The funds eventually end up at Source B's address, but the smart contract's logic obscures the link between A and B.
- **Risk**: Smart contracts are traceable on-chain unless additional privacy techniques (like zero-knowledge proofs) are implemented.

1.3.8 Multiple Wallets (Address Hopping).

- Source A can create multiple wallet addresses and send funds across many wallets before they reach Bob (Source B). This technique is sometimes called "address hopping."
- **Risk**: It's still possible to trace these transactions if someone performs **blockchain forensics**.

1.3.9 Using NFTs and Virtual Assets.

- Another emerging method involves using NFTs (non-fungible tokens) or virtual assets in the **Metaverse** platforms.
- **How it works**:
  1. Source A buys an NFT from Source B (or indirectly from Source B via an intermediary).
  2. The funds are disguised as legitimate transactions for digital assets.
  3. Since NFTs can be sold at arbitrary prices, Source B could "sell" an NFT at a higher value than its real worth to Source A, hiding the movement of funds.

2 Data Analysis

2.1 Research Questions

The following research questions were devised in analyzing the Decentraland trades.

RQ1 What are the peak periods of activity for buying and selling parcels in Decentraland?
RQ2 Who are the most influential buyers and sellers in the Decentraland marketplace?
RQ3 How do fluctuations in gas fees impact the volume of transactions in Decentraland?
RQ4 Are there instances of parcels being resold, if so, what are the profits and holding periods associated with these transactions?

2.2 Dataset

The primary dataset used is the IITP-VDLand dataset [4], a comprehensive collection of data on Decentraland parcels. As the dataset transactions were extracted from OpenSea API, the initial inspection found no additional data cleaning is required to be performed, as they were comprehensive for our analysis. It includes a rich array of attributes encompassing the following segments:

1. **Characteristics Data**: characteristics attributes of virtual parcels such as parcel ID, coordinates, owner ID, etc.
2. **OpenSea Bidding Data**: bidding (offers) proposed by potential buyers, such as parcel ID, timestamp, amount, payment details, etc.
3. **OpenSea Sales Data**: historical sales details of parcels that are sold, such as token ID, timestamp, buyer and seller address, transaction hash, etc.
4. **Ethereum Transactions Data**: transactions data associated with the parcels, such as transaction hash, buyer and seller addresses, block timestamp, payment details, etc.
5. **Social Media Data**: not used, though social media platforms were found to be influential in changing the pricing dynamics of the parcels, initial inspection of the data was found to serve little analysis for this research.

The data contains the records from March 19, 2018, to November 03, 2023, with the following statistics.

Type	Figure
Parcels Count	92,598
Number of Bidding Data	2,246,546
Number of Sales Data	20,091
Number of Ethereum Data	60,287

The majority of the research questions were analyzed with data parsing over the dataset using pandas, numpy, networkx, and seaborn, with the following section describing the details of the analysis.

2.3 Results & Discussion

In this section, we provide the results of our analysis and attempt to answer our research questions.

RQ1: What are the peak periods of activity of buying and selling parcels in Decentraland?

The BiddingData and SalesData were analyzed, and the hours of the occurrences of the transactions were aggregated with the 'event_timestamp' attribute and standardized to the UTC timezone. Figure 1 and 2 showed the hourly transaction volume for the number of biddings and the number of successful sales respectively.

The Decentraland parcel transactions reveal distinct patterns in bidding activities across different hours of the day. From Figure 1, it is observed that the bidding activity reaches its peak periods during the early hours of the day (between 4:00 AM to 8:00 AM UTC). This pattern suggests a high user engagement which could be attributed to global participation of higher internet usage times across different time zones [7, 17]. Due to the decentralized nature of transactions and the lack of timezone data, we could not determine the exact time zones with higher activity rates.

Additionally, Figure 2 shows the number of successful transactions (sales) per hour, revealing that transaction volumes are relatively low during the early morning hours (1:00 AM to 5:00 AM) but increases significantly starting from noon with a peak volume at 7:00 PM. This trend aligns with the after-work or leisure periods when users are more likely to engage in online activities, including trading virtual assets like NFT [17]. This combination of higher bidding volumes in the early hours and higher sales volumes in the evening

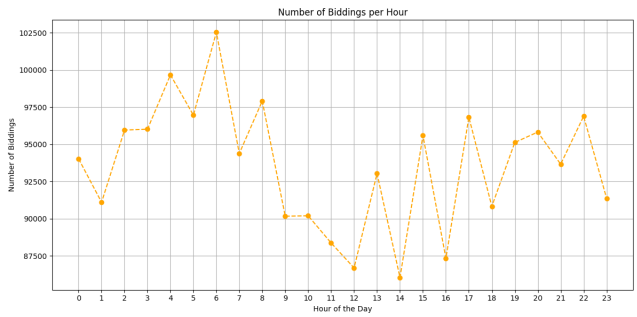


Figure 1. Number of Biddings Per Hour

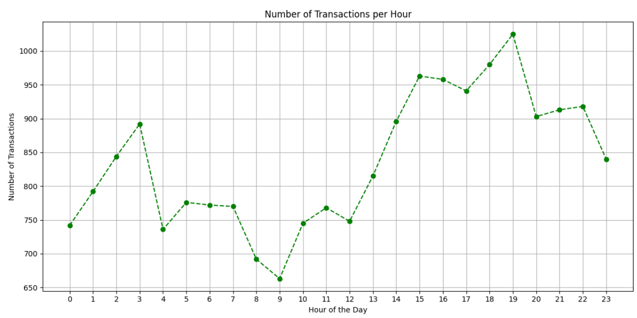
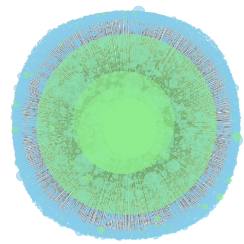
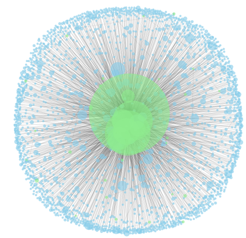


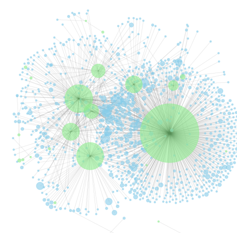
Figure 2. Number of Sales Per Hour



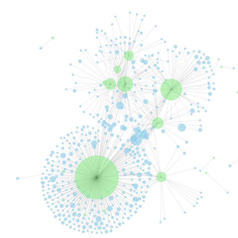
(a) All (20,091) transactions



(b) 4,000 transactions



(c) 2,000 transactions



(d) 1,000 transactions

Figure 3. Transactions Network for Buyers and Sellers

(late) hours indicates a possible competitive bidding environment, similar to a timed-auction style where higher bids are closed and finalized at the end of the day.

Another possible explanation for these observations is the influence of time zone differences and global participation in the Decentraland marketplace. Users from various parts of the world contribute to the trading volume at different times with various peaks. These patterns are also consistent with the findings in other virtual markets, such as the Bitcoin transaction volume illustrated by Dirk et al. [3] and the waves of transactions in NFT market by Christopher et al. [9]

RQ2: Who are the most influential buyers and sellers in the Decentraland marketplace?

The SalesData and OffersData of about 2,246,545 and 20,091 of transactions respectively were used to analyze this. Out of the 20,091 successful transactions, there were about 8,279 unique buyers and 99 unique sellers, bringing an average ratio of about 83 buyers per seller.

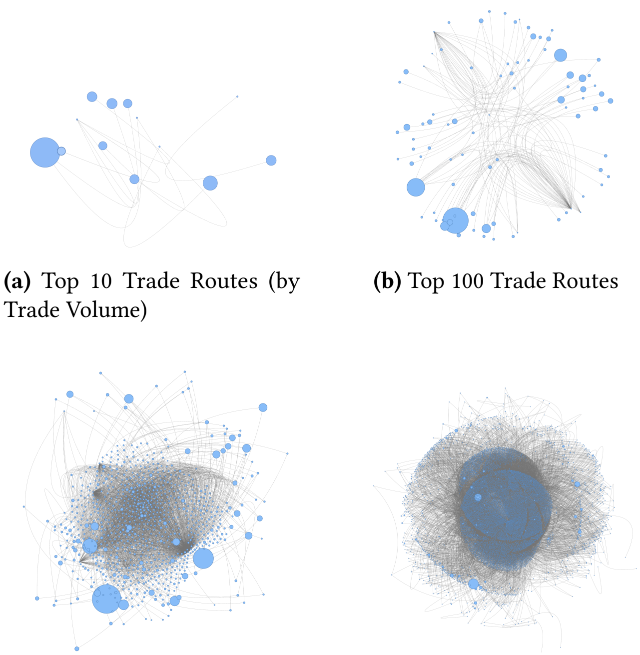
The data was graphed and Figure 3 shows the trading network between the buyers and sellers. The green circles represent sellers, while the blue circles represent buyers. The size of each bubble corresponds to the number of transactions, and the arrows indicate transactions between specific buyers and sellers.

It highlights that a few sellers represent a large portion of the transaction volumes, indicated by the size of the larger bubbles along with substantial buyers associated with it. This pattern suggests that the Decentraland marketplace is highly centralized, with a handful of sellers driving the majority of trading activities. This concentration of activity among a few key players is not unique to Decentraland but a common pattern in the digital marketplace of NFTs [1].

Moreover, the presence of such a highly active core group of sellers (and buyers) suggest that the Decentraland market may be subject to speculative trading behaviours, where the key players could create volatility in the prices, as their large transactions can shift supply and demand dynamics in the market (i.e., controls are in the hands of a few sellers). One study found traders (sellers) employed automated, high-frequency, advanced NFT strategies and often deceptive to disrupt prices and extract higher profits from a 'fair market' [6]. In other words, despite Decentraland being a fully distributed metaverse, the influence of a handful of sellers could potentially create a 'monopoly' of transaction volumes and prices [19].

Several other studies have also found similar patterns and concentrations such as herding (grouping) of NFT assets across various markets including Decentraland [2], and the NFT trade patterns by Matthieu et al. [15].

Figure 4 provides an extended overview of the trading network structure. The EthereumData transactions were used to graph the network. Each edge in the graph represents a Trade Route, the route a trade has taken from one user to another user, with propagating effects to another user(s) (i.e,



(a) Top 10 Trade Routes (by Trade Volume) **(b)** Top 100 Trade Routes

(c) Top 1,000 Trade Routes **(d)** The Whole Network

Figure 4. Trade Routes (Edges) Between Ethereum Addresses (Nodes) Simplified As An Unweighted Undirected Graph: Node Size Indicates Trade Volume (Number Of Trades)

A buys from B, B buys from C, C buys from D, etc., with the same parcel propagates D->C->B->A). Each node represents a user (buyer/seller) with its size indicating the number of trades/transactions.

Combining the analysis from both Figure 3 and 4, the same parcel gets passed through from one user to another user with only a dominant number of users running all trades. This sort of exhibits some arbitraging patterns, with further analysis to be discussed in the RQ4 in section 2.3.

RQ3: How do fluctuations in gas fees impact the volume of transactions in Decentraland?

The EthereumData were analyzed. Each transaction data associated with a parcel of bids and sales comes with a transaction hash, its value, and the gas price charged in the transaction. Gas is the fee required to successfully execute a contract on the Ethereum blockchain. The EthereumData was parsed and filtered out for unsuccessful transactions, and the gas price was averaged for each day between 2018 and 2021.

Figure 5 shows a significant relationship between the average gas price (in Gwei) and the number of transactions on the Decentraland platform over time. Gwei is a denomination of the cryptocurrency Ethereum (ETH), it is used to measure the amount of computational effort required to execute transactions on the Ethereum network. 1 Gwei is equivalent to 1-billionth of 1 ETH (or 0.000000001). Figure 5 demonstrates an inverse correlation between gas fees and transaction volumes. During periods of high gas prices, such as the peaks observed in late 2020 and early 2021, the number of transactions tends to decrease. Conversely, when gas volumes are lower, transaction volumes generally are higher. This pattern suggests that higher transaction costs deter users from engaging in the marketplace, a phenomenon consistent with the economic principles of supply and demand where higher costs reduce activity.

Initial launch and crypto bubble. Notable periods, such as the initial launch of Decentraland in 2017/2018 and the 2020-2021 cryptocurrency bubble, were also highlighted in Figure 5. During the initial release period in late 2017 and early 2018 (though not made to the public until 2020), the diagram shows a substantial spike in number of transactions. This can be attributed to the high demand for Ethereum network resources as new users and investors flocked to acquire parcels in Decentraland, causing a surge in transaction volume.

Another observation is that the cryptocurrency market experienced a significant bubble in 2020-2021, characterized by a dramatic drop in the value of various cryptocurrencies, including Ethereum plummeted more than 40%, causing heightened gas prices and reduced trading activity across platforms with heavy ETH transactions such as Decentraland.

Indirect influence from other transactions. The Decentraland transactions are processed on the Ethereum network (for ETH-based payments), it shares block spaces with all other ETH-based transactions. Several studies have found propagating effects of transactions volume across ETH network [8, 12] caused by high gas price increase of an activity in other areas of the ecosystem (e.g., NFT drop, DeFi boom).

Gas price distribution. Figure 6 shows (an enlarged version) the distribution of gas prices, indicating the median and mean gas price to be about 30 and 55 Gwei respectively. These could be attributed to the economic burden imposed by high gas fees, which can significantly increase the overall cost of transactions. Abdul & Samir [11] observed similar behaviours in trading activity in blockchain environment, where gas fees significantly increase the overall cost according to platform usage, influenced by the activity or conditions of the user. Although the median of 30 Gwei (about USD3.74 is relatively low in comparison to the overall price of a Decentraland parcel), a study has shown that gas fees require careful management as high fees could affect the scalability of a blockchain and result in the use of alternative blockchain, leading to a fragmentation of the crypto landscape [5].

Moreover, the fluctuating nature of gas fees, driven by the demand for Ethereum network resources, introduces an element of unpredictability that can discourage frequent

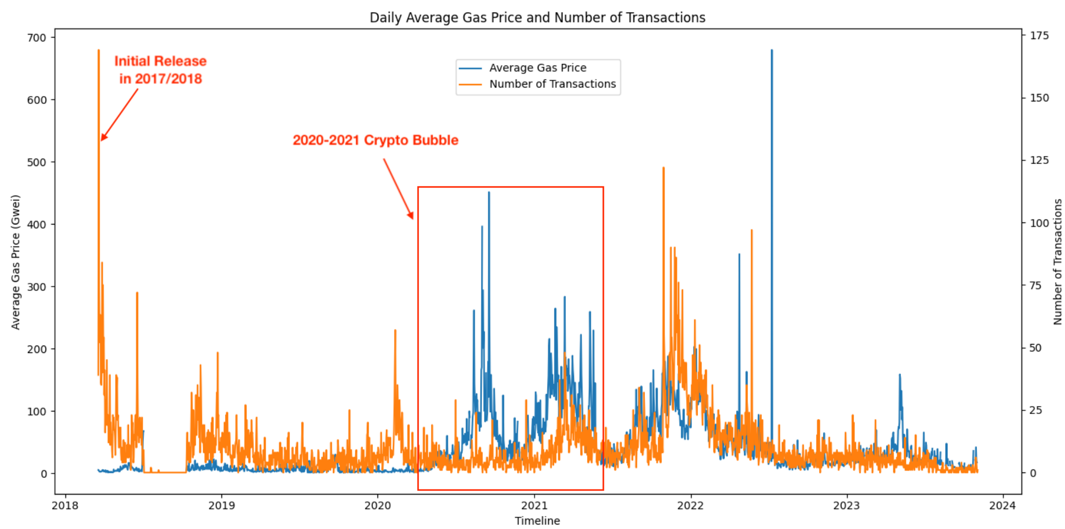


Figure 5. Gas Timeline

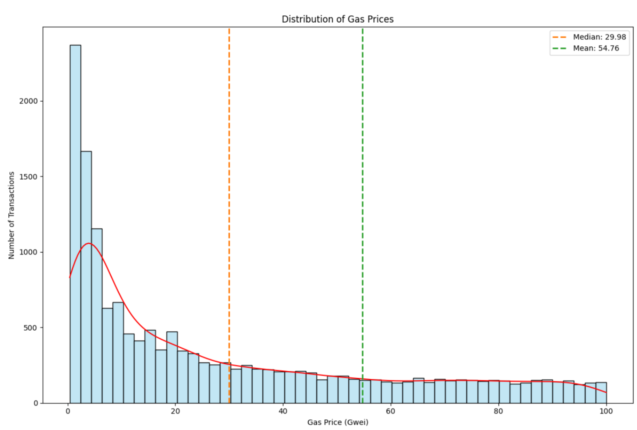


Figure 6. Gas Price Distribution

Type	# of Users
Buyers	8,142
Sellers	4,358
Total (Buyers & Sellers)	9,831
Arbitrageurs	2,669

Table 1. Number of Users by Type

Type	Mean	Median	Max	Min
Profits	-$12.7	$278.5	-$118,672	$132,778
Days Arbitrage	149	18	0	2048

Table 2. Statistics for Profits and Days Arbitrage

trading. When gas prices are volatile, users may opt to minimize transactions to avoid unexpectedly high fees, a behavior documented in studies of other cryptocurrency markets [5].

RQ4: Are there instances of parcels being resold, if so, what are the profits and holding periods associated with these transactions?

The first step of the analysis for this RQ involved extracting the buyer and seller addresses from the sales transactions data, followed by filtering the addresses that appeared both as buyers and sellers; this helps pinpoint users who engaged in the resale of parcels, thereby isolating potential arbitrage activities. Then, the transactions of the same parcel being bought and sold by the potential arbitrageurs are filtered (with the 'token_id'). An additional condition of selling date >= buying date is also filtered as one could not sell before one buys/owns. The following data were obtained.

Out of 9,831 total users, 2,669 engaged in arbitrage activities, amounting to approximately 27.15%. This substantial participation rate underscores a vibrant market within Decentraland, where users actively seek to capitalize on price fluctuations to generate profits.

The analysis of resale activity within Decentraland is as depicted in the Figure 7. The scatter plot, which correlates profits with days arbitraged, indicates a wide range of outcomes for users engaging in buying and selling of parcels. Note that the figure is scaled in with outliers data not displayed.

Notably, while the mean profit from resales is approximately a loss of $12.70, the median profit stands at $278.46, suggesting that although the average user might experience a loss, a substantial number of users can achieve significant gains. Furthermore, there are profits as high as $132K and as low as $118K. This discrepancy highlights the variability and

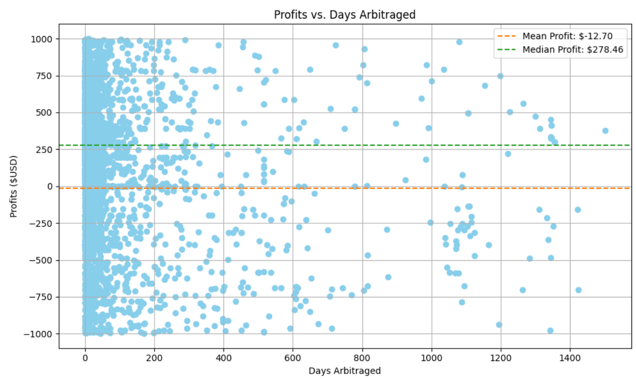


Figure 7. Scatters of Profits and Days

speculative nature of the Decentraland market, where successful arbitrage can yield high returns, but risks are equally high.

The median and mean holding period of the parcels are 18 days and 149 days respectively, suggesting that most profitable trades are executed within a relatively short timeframe, aligning with speculative trading behavior where quick turnarounds are common [2, 6]

These resales of parcels are conducted in a 'secondary market' where people buy/sell assets (or virtual lands) that have already been created and sold previously. Studies conducted by Seboem et al. [16] found that a secondary market on OpenSea amounted to as high as 46%, indicating a high profitability from trade arbitraging.

3 Data Driven Pattern Recognition

In this section, we delve into the second major contribution of this paper, which involves applying the GTN2vec [14] argorithm to the IITP-VDLand dataset [4] and use its graph-embeddings to train some initial models and discover potential anomaly. GTN2vec is a graph embedding algorithm specifically designed for detecting money laundering activities on the Ethereum blockchain. This algorithm is innovative in its use of transaction-specific features, such as gas price and timestamps, to enhance the feature extraction process and improve the detection of illicit accounts.

3.1 Model Overview

GTN2vec is a graph embedding algorithm specifically designed for detecting money laundering activities on the Ethereum blockchain. The algorithm transforms Ethereum transaction data into a directed, weighted graph where nodes represent Ethereum accounts, and edges signify transactions between them. By incorporating transaction-specific features, such as gas price and timestamp, into the graph's edge weights, GTN2vec captures behavioral patterns that are unique to money laundering activities.

The *gas price* reflects the urgency of a transaction and its likelihood of being prioritized by the network, which is often manipulated by money launderers to expedite fund transfers.

The *timestamp* provides temporal information about transaction sequences, highlighting rapid transfers that are typical in money laundering schemes. By fusing these attributes with the structural information of the transaction graph, GTN2vec generates low-dimensional embeddings that effectively represent the characteristics of each node (Ethereum account).

The key innovation of GTN2vec lies in its weighted random walk mechanism. Transition probabilities are calculated using a combination of gas price and timestamp, controlled by a balance parameter α. This ensures both features contribute effectively to the embedding process:

$$Pwt_{ux} = Pw_{ux}^{\alpha} \cdot Pt_{ux}^{1-\alpha}$$

where Pwt_{ux} is the combined weight, Pw_{ux} represents the normalized gas price, and Pt_{ux} denotes the normalized timestamp.

Furthermore, GTN2vec introduces customizable parameters to enhance the random walk process:

- **Return Parameter (p)**: Adjusts the probability of revisiting the previous node.
- **Exploration Parameter (q)**: Balances between depth-first (local) and breadth-first (global) exploration.
- **Balance Parameter (α)**: Controls the relative importance of gas price and timestamp.

These features enable GTN2vec to capture both local and global structures of the transaction graph while emphasizing transaction-specific behaviors that are critical for detecting illicit activities.

4 Implementation

The implementation of the methodology consisted of several sequential steps, beginning with data preprocessing and culminating in the application of various labeling and classification strategies. Below, we detail each step of the process.

4.1 Dataset Description and Preprocessing

The dataset used for this project was the IITP_VDLand dataset, which contains transaction records from Decentraland on the OpenSea platform. This dataset was derived from on-chain Ethereum data and served as the basis for constructing a graph representation of transaction activity.

Initially, the dataset underwent a preprocessing phase to prepare it for analysis. The preprocessing involved filtering the dataset to retain only the required columns and removing duplicate entries to ensure data consistency. Additionally, normalization was applied to certain numeric columns, such as `gas_price` and `transaction_timestamp`, to scale their values and improve comparability.

4.2 Graph Construction

Using the cleaned dataset, a graph representation was constructed with the `networkx` library. In this graph, each node represented an Ethereum wallet address, and each edge corresponded to a transaction between two addresses. The edges were weighted based on transaction metadata, such as transaction value and gas price. This graph served as the foundational structure for further analysis, including the generation of node embeddings.

4.3 GTN2Vec Algorithm

To extract meaningful embeddings from the graph, the GTN2Vec algorithm was implemented, following the methodology described in the reference paper. The implementation involved the following steps:

1. **Define Transition Probabilities:** Transition probabilities were calculated to guide biased random walks on the graph. These probabilities determined the likelihood of transitioning between nodes during the walk, based on their connectivity and edge weights.
2. **Perform Biased Random Walks:** Simulated biased random walks on the graph to generate sequences of nodes. These sequences captured the structural relationships between nodes within the graph.
3. **Generate Random Walks:** Multiple random walks were generated for each node in the graph to ensure sufficient coverage of the graph's structure.
4. **Convert Walks to Node Embeddings:** The node sequences generated from the random walks were used to produce vector embeddings for each node. These embeddings encoded the structural and transactional relationships present in the graph.

4.4 Labeling Strategies

One of the significant challenges in this project was the absence of labels for money laundering within the IITP_VDLand dataset. To address this, three labeling approaches were explored.

Approach #1: Use Labeled Sources (Failed). This approach attempted to integrate labeled Ethereum transaction data from external sources. Two labeled datasets from Kaggle were identified, containing a total of 5,390 flagged suspicious wallets or contracts.

These flagged accounts were merged into a single list and compared against the IITP_VDLand dataset. However, there was no overlap between the flagged addresses in the Kaggle datasets and the IITP_VDLand dataset, rendering this approach unsuccessful.

Approach #2: Heuristic-Based Labeling. In the second approach, heuristic rules were defined to label transactions within the IITP_VDLand dataset. The heuristics included the following criteria:

- **Big Transactions:** Transactions with a `value_in_wei` exceeding a predefined threshold (e.g., 10^{18} wei, equivalent to 1 ETH) were flagged as suspicious.
- **Frequent Transactions:** Addresses with a high number of transactions within a specified timeframe (e.g., a single day) were flagged.
- **Gas Price Anomalies:** Transactions with gas prices exceeding a high percentile threshold (e.g., the 99th percentile) were flagged.

Following the application of these heuristics, transaction-level flags were aggregated to label nodes within the graph. For example, the average flag values of transactions associated with a node were used to determine whether the node should be flagged as suspicious. While this approach resulted in labeled data, the performance of models trained on this data was moderate, with noticeable room for improvement in precision and recall.

Approach #3: Unsupervised Learning. As an alternative to labeling, unsupervised learning was explored using clustering techniques. Node embeddings were clustered using the KMeans algorithm, and dimensionality reduction via Principal Component Analysis (PCA) was employed to visualize the clustering results. Despite these efforts, the unsupervised approach was ineffective, as it failed to produce distinct clusters or clear patterns within the data. This highlighted the limitations of clustering for this task in the absence of domain-specific enhancements or labeled data.

4.5 Classification

Using the labeled data from the heuristic-based approach, a classification model was trained to distinguish between dodgy (suspicious) and non-dodgy nodes. The model demonstrated moderate success, indicating a capacity to identify suspicious nodes, but struggled to achieve a balance between precision and recall. This outcome suggested the need for further refinement, either through improved labeling techniques, additional features, or alternative modeling approaches.

4.6 Summary of Results and Challenges

The implementation successfully demonstrated the feasibility of graph-based approaches to transaction analysis. However, the lack of labeled data and the challenges associated with unsupervised learning underscored the need for further investigation. Future work could involve leveraging advanced graph neural network techniques or sourcing additional labeled datasets to enhance the robustness and reliability of the analysis.

5 Threats to Validity

While this study provides significant insights into the virtual land market on Decentraland, several potential threats to the validity of our findings need to be addressed.

5.1 Analysis and Visualization

Data Accuracy: The reliability of our interpretations heavily depends on the accuracy of the data sourced from the IITP-VDLand dataset and the OpenSea API. Any errors in data collection, recording, or extraction could introduce inaccuracies. Although efforts were made to clean and analyze the data, there could be possible inaccuracies such as incomplete transaction records, misreported prices, or incorrect timestamps that could skew the analysis.

Selection Bias: Our analysis might be influenced by selection bias if the dataset does not accurately represent the entire Decentraland market. For example, if the dataset primarily includes transactions from highly active users or specific types of parcels, the findings may not apply to the broader market. Additionally, if the dataset excludes private sales or transactions on other marketplaces, our analysis could overlook significant portions of the market.

Temporal Validity: Our study covers transactions from March 19, 2018, to November 3, 2023. The relevance of our findings might diminish over time as market conditions, user behaviors, and technological developments evolve. For example, future changes in Ethereum's transaction fee structure, the introduction of new virtual worlds, or shifts in user preferences could alter the dynamics of the virtual land market, making our conclusions less relevant.

generalizability: While our study provides insights into Decentraland, the findings may not be generalizable to other virtual worlds or NFT markets. Differences in platform design, user demographics, and market structure can lead to varying economic behaviors, limiting the applicability of our conclusions to other contexts.

5.2 Implementation

While the implementation and experiments described in this work aim to provide insights into identifying suspicious activities in Ethereum transactions, several limitations and threats to validity could affect the generalizability and reliability of the results. These are detailed below:

Dataset Limitations. The dataset used, IITP_VDLand, was derived from Decentraland transactions on the OpenSea platform. However, this dataset lacks ground truth labels for suspicious or non-suspicious transactions. This limitation necessitated the use of heuristic-based labeling, which may not fully capture the nuanced behaviors of illicit activities. Additionally, the absence of overlap between IITP_VDLand and other labeled datasets (e.g., Kaggle datasets) restricted the possibility of cross-validation with existing suspicious account lists.

Heuristic-Based Labeling. The labeling approach relied on manually defined heuristics, such as thresholds for high-value transactions, frequent transactions, and gas price anomalies. While these rules were informed by plausible patterns of suspicious behavior, they may introduce biases or fail to account for more sophisticated money laundering techniques. Consequently, the labels generated may not accurately reflect real-world scenarios, potentially impacting the performance of the classifier.

Unsupervised Learning Assumptions. The unsupervised clustering approach assumes that similar node embeddings correspond to similar behavioral patterns. However, in the absence of explicit labels or domain-specific validation, this assumption remains unverified. The lack of distinct clusters in the results further highlights the difficulty of identifying meaningful patterns using unsupervised methods alone.

Model Generalizability. The trained classifier exhibited moderate performance, but its ability to generalize to other blockchain transaction datasets or real-world scenarios remains uncertain. The model's reliance on specific features and heuristics may limit its applicability to datasets with different structures or distributions.

Graph Representation Limitations. The construction of the graph relies on the assumption that relationships between addresses can be effectively captured through transactions. However, this representation may not fully account for complex, multi-layered interactions in the Ethereum ecosystem, such as interactions through intermediary contracts or mixers. These interactions could obscure patterns of suspicious behavior, leading to potential misclassification.

Evaluation Metrics. While metrics such as precision, recall, and silhouette score provide an indication of the model's performance, they may not comprehensively reflect its effectiveness in real-world scenarios. The evaluation did not account for the potential consequences of false positives or false negatives, which could have significant implications in practical applications.

Domain-Specific Knowledge. The lack of domain-specific knowledge in defining features or constructing heuristics poses another threat to validity. Money laundering techniques often evolve rapidly, and without expert input, the approach may fail to adapt to emerging patterns of suspicious behavior.

Ethical and Practical Constraints. Finally, ethical and practical constraints in accessing labeled datasets, such as those maintained by financial institutions or law enforcement agencies, limit the ability to validate the model on real-world data. Additionally, the reliance on publicly available datasets may introduce biases that do not generalize to broader transaction datasets.

6 Future Work

Our research on the economics of virtual worlds, specifically within Decentraland, has unveiled numerous insights and patterns. However, several areas warrant further exploration to deepen our understanding and enhance the robustness of our findings. Future work could focus on the following aspects:

6.1 Analysis

Advanced Methodologies: Implementing more sophisticated analytical techniques, such as machine learning algorithms and advanced statistical models, could provide deeper insights into NFT market dynamics and user behaviors. Predictive modeling, anomaly detection, and clustering algorithms could help identify patterns and trends not captured by our current methods.

Social Media Integration: Incorporating data from social media platforms for sentiment analysis could offer valuable insights into how public perception and discussions influence trading behaviors and market trends. Machine learning models can analyze large volumes of social media data to detect sentiment shifts that correlate with market activities.

Cross-Market Comparisons: Conducting comparative studies between Decentraland and other virtual worlds or NFT marketplaces could reveal unique and shared economic dynamics. Such comparisons would help identify best practices, potential vulnerabilities, and areas for improvement across different platforms.

User Behavior Analysis: Delving deeper into user behavior, such as investment strategies, holding periods, and transaction patterns, could provide a granular understanding of market participants. Segmenting users based on their trading activities and analyzing their motivations and decision-making processes would offer valuable insights for both researchers and market participants.

6.2 Implementation

The findings and limitations of this work suggest several avenues for future research and development. These directions aim to address current challenges and improve the robustness, accuracy, and applicability of the methods employed.

Improved Labeling Techniques. One of the key challenges in this work was the lack of labeled data for the IITP_VDLand dataset. Future efforts could focus on obtaining labeled datasets, either through partnerships with financial institutions, law enforcement agencies, or blockchain analytics firms. Additionally, incorporating semi-supervised learning methods could leverage both labeled and unlabeled data to improve labeling accuracy.

Advanced Graph Representation. The graph representation used in this study primarily modeled direct transactions between wallet addresses. Future work could explore more sophisticated graph representations that incorporate additional layers of interaction, such as intermediary contracts, token-specific transactions, and temporal dynamics. Multi-relational or heterogeneous graphs could provide deeper insights into complex blockchain interactions.

Graph Neural Networks. Graph Neural Networks (GNNs), such as GraphSAGE, Graph Attention Networks (GATs), or Graph Convolutional Networks (GCNs), offer promising alternatives for learning node embeddings. These models can capture both local and global graph structures and could improve the detection of suspicious patterns in the transaction graph.

Enhanced Feature Engineering. Future research could incorporate domain-specific features, such as transaction volume trends, token types, smart contract interactions, and temporal patterns. Feature selection techniques could also be applied to identify the most relevant attributes, reducing noise and improving model performance.

Exploration of Alternative Clustering Methods. Given the limited success of KMeans clustering, alternative unsupervised learning approaches, such as DBSCAN, hierarchical clustering, or spectral clustering, could be explored. These methods may better handle noise and irregular cluster shapes, potentially improving the identification of suspicious behaviors.

Ethical Considerations and Collaboration. Future work should also address ethical concerns associated with the use of blockchain data for monitoring suspicious activities. Collaboration with legal and regulatory bodies could ensure that proposed methods align with ethical and legal standards.

Cross-Blockchain Analysis. Money laundering schemes often span multiple blockchain networks. Future research could investigate cross-blockchain analysis, integrating data from various blockchain platforms to identify suspicious activity that might be missed when analyzing a single blockchain in isolation.

7 Source Code Availability

To ensure transparency, the source code used for data analysis and visualization is publicly available. The repository includes scripts for parsing and analyzing the IITP-VDLand dataset, generating the figures presented in this paper, implementing the GTN2Vec algorithm, and demonstrating its application in creating both supervised and unsupervised data-driven models.

The repository is hosted on GitHub and can be accessed via the following link: GitHub Repository

8 Conclusion

This study delved into the economics of virtual worlds, with a specific focus on the virtual land market within Decentraland as facilitated through transactions on OpenSea. By leveraging the comprehensive IITP-VDLand dataset, we were able to address several key research questions that shed light on the dynamics of this nascent market.

Our analysis uncovered distinct patterns in trading activities, revealing peak periods for buying and selling, and identifying key players who significantly influence market dynamics. We found that fluctuations in Ethereum gas fees have a substantial impact on transaction volumes, with higher fees deterring market participation. Additionally, our investigation into resale activities highlighted the speculative nature of the market, with a notable portion of users engaging in arbitrage to capitalize on price differentials.

These findings underscore the complexity and volatility inherent in virtual asset markets. The presence of a few dominant sellers and the influence of external factors such as gas fees and broader market trends highlight the need for more robust regulatory frameworks to mitigate risks and protect participants. Moreover, the high level of activity from a relatively small group of influential players suggests potential vulnerabilities to market manipulation.

The study's limitations, including potential biases in the dataset and the evolving nature of the virtual land market, suggest that our conclusions should be interpreted with caution. Nonetheless, our research provides a foundational understanding of the economic behaviors within Decentraland, offering valuable insights for stakeholders and policymakers.

Future research could expand upon this foundation by exploring cross-platform trading activities to examine the interconnectedness of virtual asset markets. Additionally, incorporating temporal and predictive modeling techniques could provide further insights into market trends and potential risks. Advancements in graph-based analysis, combined with machine learning approaches, may enable deeper exploration of user behaviors and the detection of fraudulent or manipulative activities. Ultimately, a more comprehensive understanding of these markets will be essential for ensuring their sustainability and protecting participants in the rapidly evolving metaverse ecosystem.

References

[1] Sajjad Alizadeh, Amin Setayesh, Arash Mohamadpour, and Behnam Bahrak. 2023. A network analysis of the non-fungible token (NFT) market: structural characteristics, evolution, and interactions. *Applied Network Science* 8, 1 (2023), 38.

[2] Te Bao, Mengzhong Ma, and Yonggang Wen. 2023. Herding in the non-fungible token (NFT) market. *Journal of Behavioral and Experimental Finance* 39 (2023), 100837.

[3] Dirk G Baur, Daniel Cahill, Keith Godfrey, and Zhangxin Frank Liu. 2019. Bitcoin time-of-day, day-of-week and month-of-year effects in returns and trading volume. *Finance Research Letters* 31 (2019), 78–92.

[4] Ankit K. Bhagat, Dipika Jha, Raju Halder, Rajendra N. Paramanik, and Chandra M. Kumar. 2024. IITP-VDLand: A Comprehensive Dataset on Decentraland Parcels. arXiv:2404.07533 [cs.LG] https://arxiv.org/abs/2404.07533

[5] Frederic Boissay, Giulio Cornelli, Sebastian Doerr, and Jon Frost. 2022. *Blockchain scalability and the fragmentation of crypto*. Technical Report. Bank for International Settlements.

[6] Priyanka Bose, Dipanjan Das, Fabio Gritti, Nicola Ruaro, Christopher Kruegel, and Giovanni Vigna. 2023. Exploiting Unfair Advantages: Investigating Opportunistic Trading in the NFT Market. *arXiv preprint arXiv:2310.06844* (2023).

[7] BYDFi. [n. d.]. Which time zones have the highest trading volume for cryptocurrencies? https://www.bydfi.com/en/questions/which-time-zones-have-the-highest-trading-volume-for-cryptocurrencies

[8] David Carl and Christian Ewerhart. 2020. Ethereum gas price statistics. *University of Zurich, Department of Economics, Working Paper* 373 (2020).

[9] J Christopher Westland. 2024. Periodicity, Elliott waves, and fractals in the NFT market. *Scientific Reports* 14, 1 (2024), 4480.

[10] DEW 2017. *Decentraland Whitepaper*. Retrieved July 14, 2024 from https://docs.decentraland.org/player/general/whitepaper/

[11] Abdul Jabbar and Samir Dani. 2020. Investigating the link between transaction and computational costs in a blockchain environment. *International Journal of Production Research* 58, 11 (2020), 3423–3436.

[12] Dimitrios Koutmos. 2023. Network activity and ethereum gas prices. *Journal of Risk and Financial Management* 16, 10 (2023), 431.

[13] Andrew Leppla, Jorge Olmos, and Jaideep Lamba. 2022. Fraud Pattern Detection for NFT Markets. *SMU Data Science Review* 6, 2 (2022), Article 21. https://scholar.smu.edu/datasciencereview/vol6/iss2/21/

[14] Jiayi Liu, Changchun Yin, Hao Wang, Xiaofei Wu, Dongwan Lan, Lu Zhou, and Chunpeng Ge. 2023. Graph Embedding-Based Money Laundering Detection for Ethereum. *Electronics* 12, 14 (Jan. 2023), 3180. https://doi.org/10.3390/electronics12143180 Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[15] Matthieu Nadini, Laura Alessandretti, Flavio Di Giacinto, Mauro Martino, Luca Maria Aiello, and Andrea Baronchelli. 2021. Mapping the NFT revolution: market trends, trade networks, and visual features. *Scientific reports* 11, 1 (2021), 20902.

[16] Sebeom Oh, Samuel Rosen, and Anthony Lee Zhang. 2022. Investor experience matters: Evidence from generative art collections on the blockchain. *Available at SSRN* 4042901 (2022).

[17] Sealaunch. 2023. Minting NFTs: what the onchain data tells us about the primary NFT market. https://sealaunch.substack.com/p/minting-nfts-what-the-onchain-data

[18] Mingxiao Song, Yunsong Liu, Agam Shah, and Sudheer Chava. 2023. Abnormal Trading Detection in the NFT Market. (2023). arXiv:2306.04643 [q-fin.TR] https://arxiv.org/abs/2306.04643

[19] Jying-Nan Wang, Yen-Hsien Lee, Hung-Chun Liu, and Yuan-Teng Hsu. 2023. Dissecting returns of non-fungible tokens (NFTs): Evidence from CryptoPunks. *The North American Journal of Economics and Finance* 65 (2023), 101892.

[20] Jiajing Wu, Kaixin Lin, Dan Lin, Ziye Zheng, Huawei Huang, and Zibin Zheng. 2022. Financial Crimes in Web3-empowered Metaverse: Taxonomy, Countermeasures, and Opportunities. http://arxiv.org/abs/2212.13452 arXiv: 2212.13452 [cs] tex.pubstate: prepublished.