Market and Network Structure Analysis of Non-Fungible Tokens

LIU Zhiruo
20884634
zliufc@connect.ust.hk

YE Yil Lin
20922644
ylye@connect.ust.hk

ZHANG Keyi
20920309
kzhangbs@connect.ust.hk

ZHANG Lu
20891297
lzhangdp@connect.ust.hk

DONG Wenyu
20930158
wdongai@connect.ust.hk

I. Introduction

NFT (Non-Fungible Token) is a cryptographic digital credential supported by blockchain technology to track the ownership of virtual digital assets such as artworks or collectibles. Based on the unique and irreplaceable properties, NFT can actualize the copyright of digital assets, facilitate the flow of digital asset transactions, and provide users with various consumption values such as collectability and investment.

***Index Terms*—NFT, market analysis, graph analysis, network structure**

A. Background and Problem Definition

With the big hit of CryptoPunks, Bored Ape Yacht Club and Decentraland, NFT draws considerable attention from the media and the general public, which also brings a series of confusion. Nonetheless, research on NFTs is limited and focused mostly on technical aspects.

In our project, we will study the characteristics of NFT market and analyze the topological structure of NFT transaction networks. Meanwhile, we compare NFT networks with other common network systems, such as social networks and computer (Web) networks.

The report is divided into three parts: data processing, characteristics of NFT market, and graph analysis of NFT transactions. Section II describes our data-collecting and cleaning process. Section III explores the NFT buyers and the characteristics of the whole market. Section IV presents the graph analysis on NFT networks: (1) defines the graph structure of NFT transactions (2) discusses in-degree, and out-degree distribution of nodes (3) analyzes the topological characteristics of NFT transaction graph (4) identifies the connectivity as well as clustering properties of such transaction graphs.

In each section, we make certain adjustments and optimizations in data, variables, and models based on studies and papers in the related domains.

B. Related Work

Based on web search analysis, which establishes graphical connection methods between nodes, more and more network analyses are used in areas such as social media and blockchain transactions. To be specific, graph analysis is commonly used to visualize the connection of data on Ether in studying blockchain transactions. The ERC20 fungible token transaction in Ether has been the subject of a few research[1], with the degree distribution of the entire ERC20 token transfer following a power law[2]. It has certain limitations on generalizability due to less pronounced heavy tails than in social networks in the network degree distribution.

II. Datasets

We selected 8 NFT projects according to market capitalization and popularity. They are Acclimated Moon Cats, Art Blocks, Bored Ape Yacht Club, Cryptopunks, CryptoVoxels, Decentraland, HahsMasks, and Meebits.

A. Data extraction and data cleaning

Alchemy platform and A16Z NFT Analyst Starter Pack were utilized to extract blockchain data by APIs it offered in order to make further analysis easier. CSV files were used to store the results. Common features of the NFTs are grouped in collections, which names are cleaned and evened out. The raw names of different NFTs, as downloaded from OpenSea, are stripped by special characters and digits, such as _, – and so on. For example, the collection ”Acclimated Moon Cats” renames all collections starting with that string of characters in Acclimated Moon Cats using regular expression. In this work, the fields include transaction hash, the NFT smart contract address, seller address, buyer address, Ethereum block timestamp, block number, the amount of ETH and/or WETH transferred, and the total number of tokens swapped in the transaction. Transactions with one of the former fields empty are deleted. Many transactions involve several addresses or transfer back to the original address, making them difficult because various records have the same transaction hash. Therefore, in this study, a simplification was implemented: the maximum amount of WETH and/or ETH sold is taken into account as a fee when the seller uses a proxy contract. Data was gathered for each NFT collection from January 1 to September 10, 2022.

B. Bulk Statistics

In this work, we defined two kinds of account and three kinds of transactions in terms of Ethereum terminology, there are two kinds of accounts:

There are 2 kinds of wallets, and the result of classification is shown in TABLE 1.

- **EoAs**: personal wallet.(i.e. to_contract is false)
- **CAs**: smart contract accounts which controlled by contract code.(i.e. to_contract is true)

TABLE I
THE RESULTS OF CLASSIFICATION ACCORDING TO WALLETS.

Collections	**Wallets**		
	Total	***EoAs***	***CAs***
Acclimated Moon Cats	2046	2037	9
Art Blocks	9591	9541	50
Bored Ape Yacht Club	7214	7185	29
Cryptopunks	4501	4010	491
CryptoVoxels	2085	2048	37
Decentraland	4321	4225	96
HashMasks	7730	7250	480
Meebits	6448	6419	29

Then transactions were divided to 3 kinds. The result of classification is shown in TABLE 2.

- **Buy/Sell**: The seller address is not equal to the buyer address and the amount of ETH and/or WETH for the token that buyer address has swapped is not 0;
- **Mint**: The seller address is started with '0x000000'.
- **Transfer**: The seller address is equal to the buyer address, and/or the amount of ETH and/or WETH that the transaction has been made is not 0;

TABLE II
THE RESULTS OF CLASSIFICATION ACCORDING TO TRANSACTIONS.

Collections	**Transactions**			
	Total	***Buy/Sell***	***Mint***	***Transfer***
Acclimated Moon Cats	14378	1739	10765	1874
Art Blocks	80660	20144	48737	11779
Bored Ape Yacht Club	32074	15872	10002	6200
Cryptopunks	28576	10459	0	18117
CryptoVoxels	16607	8394	5123	3090
Decentraland	174322	540	57631	116151
HashMasks	49404	15624	16385	17395
Meebits	29224	6250	20000	2974

Statistics data for the classification are summarized in Table 3.

TABLE III
DESCRIPTION ABOUT ETH/WETH.

Collections	**ETH / WETH(highest)**			
	Total	***Max***	***Average***	***Std***
Acclimated Moon Cats	701.15	29.00	0.1469	0.6979
Art Blocks	21725.59	65.00	0.0387	0.4949
Bored Ape Yacht Club	31487.72	109.00	0.1832	2.5492
Cryptopunks	154139.63	4200.00	5.3966	29.9446
CryptoVoxels	16960.02	250.00	1.0940	4.2372
Decentraland	1318.97	78.93	0.0433	0.7252
HashMasks	120163.19	420.00	2.432	5.717
Meebits	43938.21	1000.00	1.503	9.230

III. CHARACTERISTICS OF THE NFT MARKET

No one knows if NFT is a scam, or a new future, a digital art, or a financial bubble. What is driving it and what factors are influencing its market movement? In 2021, approximately 11 million art sales took place, driving the booming NFT market[8]. People may have questions, is that a small group of people each holding large amounts of NFT? Are they driving the hype and carrying the market? In order to figure out these mysteries, we focused on the data itself at first to learn more about the characteristics of the NFT market.

We analyzed the number of tokens owned and the corresponding number of addresses owning tokens by using the histogram and can be seen from Figure 1 that the vast majority of NFT owners each only owned a small number of tokens, with few addresses owning hundreds or even thousands of tokens. To better understand the dataset, we then

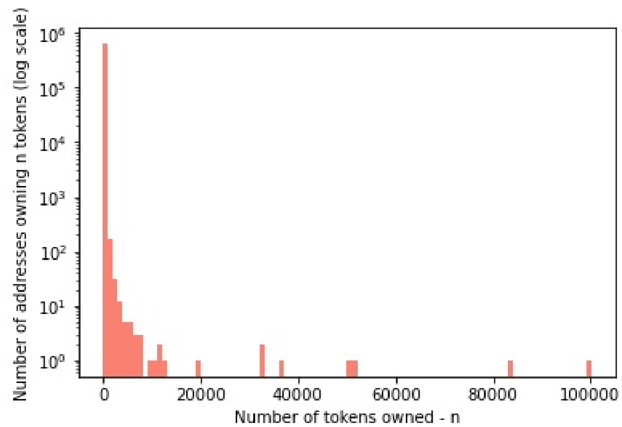

Fig. 1. Distribution of the Whole Dataset

analyzed the owners without many tokens. Figure 2 shows the data with a logarithmic scale. We treated log of rank of token owners be the rank, and log of number of tokens be the frequency. It is obvious to see that the value on the x-axis has an inverse relation with the value on the y-axis. So, we deduce the data follows a Zipf Law Distribution[11]. That is, the higher the rank of the token owner, the lower the number of tokens owned. To confirm this finding, we chose a cut-off point, n=2000, for the number of tokens owned to help evaluate the trend of NFT ownership in more detail. The results are presented in Figure 3. With these findings, we draw the conclusion that the data indeed follows a Zipf law distribution[11], and the NFT market is open and decentralized. Since it is no longer controlled by any central entity, there are fewer barriers to entry for buyers who want to enter this market. This is why we are able to see in the data distribution that there are far fewer large-scale NFT owners on the Ethereum blockchain. Today, almost all games contain their own virtual goods. As a business model, it has even spawned a black market. Digital goods, collectibles and assets are no longer just for the geeks and elites, they are

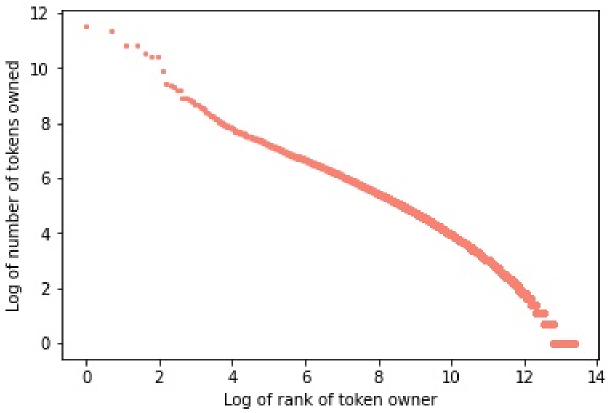

Fig. 2. Variables Comparison on a log-log Scale

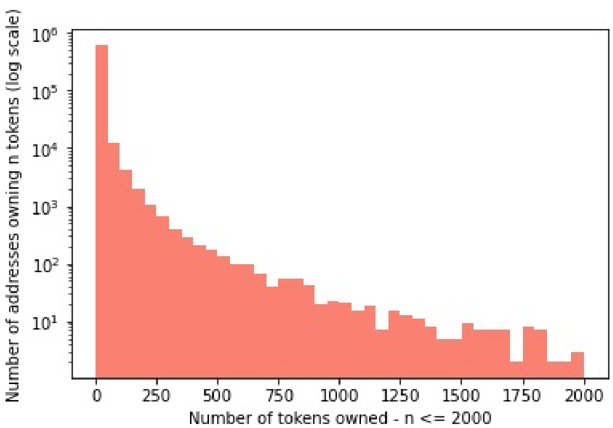

Fig. 3. Zoom in the Data Distribution with n = 2000

also profitable investments. So, the NFT market exists and thrives for a reason.

Addresses as our main object of study cannot be equal to real-world entities. An entity may manage NFT through a different address. Although this is technically difficult to achieve, it cannot be ruled out that some players operate in this way.

IV. ANALYSIS OF NETWORK STRUCTURE

In this part, we study the structural characteristics of each NFT project networks by measuring their statistical properties. To be specific, we discuss the diameter, average shortest path length, the distribution and density of in and out degree, and the connectivity and clustering properties. In addition, we compare the qualities of the NFT networks with those previously noted for the Web and social networks.

A. Define Graph

With a group of nodes V and edges $E = \{(v_i, v_j)^k | v_i \in V, v_j \in V, k \in N\}$, the transactions can be structured in a Multi-directed Weighted Graph. Each directed edge $e^k_{i,j}$ represents a single transaction from v_i to v_j, and each node v_i denotes a distinct wallet. Additional data is defined and saved for each transaction as an edge coefficient, including the weight of the edge - $Cost(v_i, v_j)^k$ equivalent to the cost paid by the wallet v_j to the wallet v_i to complete the transaction, the identity of the transaction record, $Transcation(v_i, v_j)^k$, and the transaction time $Time(v_i, v_j)^k$ [4].

B. Setting Up for Graph Analysis

In this paper, we only consider the transactions involving two EoA addresses since the primary goal of our research is to learn about and examine how groups/ communities that engage in the purchasing and selling of NFTs behave. Therefore, all transactions done through smart contracts CAs rather than through generic wallets EoA were dropped from the graph in an effort to at least partially reduce the biases. Moreover, the node involving mint was also removed.

C. Node In-degree and Out-degree Distribution

To start, we examine the graph topology by studying the distribution of node degrees. It allows us to understand the user behaviors of the transaction of a specific NFT project.

It has been shown that many complex networks, including online social networks, have degree distributions that follow power laws [1]. Therefore, the fact that social networks likewise have power-law degree (highly skewed, heavy-tailed degree distribution) distributions may not come as a surprise.

The following formula describes the power law distribution.

$$p(x) = Cx^{-a}$$

Figure 4, 5 display the in-degree and out-degree complementary cumulative distribution functions (CCDF) for each NFT collection that was measured. Across all of the networks, the majority of the nodes have a low degree, whereas a tiny number of nodes have a noticeably greater degree. This behavior is consistent with a power-law network. We determined the optimal power-law coefficient using the maximum likelihood approach and evaluated how well the degree distributions are described by a power-law [2]. The computed power-law coefficient and the Kolmogorov-Smirnov goodness-of-fit measure are displayed in Table IV. We can see that all the NFT collection graphs are well approximated by a power-law.

Figure 7 visualizes the degree of CrytoVoxels, Meetbits, and combined transaction network with probability density functions on Log-Log axes which allows us to better visualized the heavy-tailed distributions. The rest of the NFT projects and also the combined NFT network are shown in figure 8. These graphs again suggest that the degree of NFT transaction graph follows the power-law distribution.

In particular, in the second row of figure 7, the green dashed line shows the Power-Law fit start from optimal Xmin whereas the dotted line starts at Xmin =1. We observe that two green lines are close to each other suggesting that every part of our data fit well on power-law. Moreover, graphs in the third row compare the goodness of fit to an exponential distribution (red dashed line), and we can clearly see that data fit much better on power law distribution.

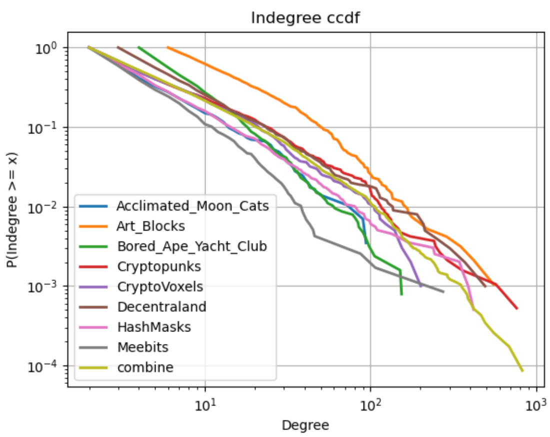

Fig. 4. In-degree complementary cumulative distribution functions (CCDF).

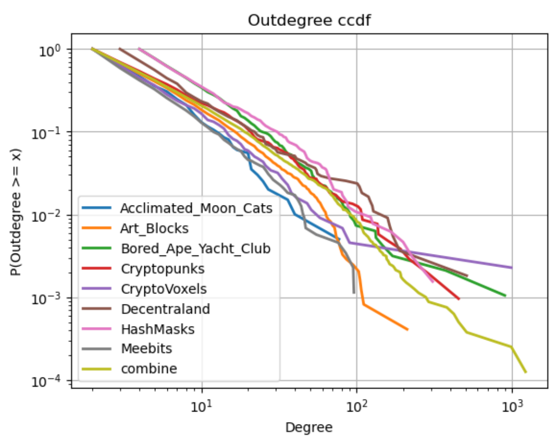

Fig. 5. Out-degree complementary cumulative distribution functions (CCDF).

This result generally shows the presence of a few high-degree hubs. Also, it implies that a significant portion of the information concerning node interactions should be drawn from a study of their tails, such as the influencers in social networks [1], and popular websites in web networks [1].

To compare the structure between NFT network, Social network and Web network, we list the power-law coefficient values for graphs created for web networks [4], social networks like Flickr, Livejournal, and YouTube, as well as their mathematical average [3].

From the table and Figure 6, we noticed that the differences of power-law coefficients of the in-degree and out-degree are small for Social Network, while the in-degree and out-degree power-law coefficients have been shown to differ significantly for both Web network and NFT network.

This suggests that, for internet social networks, the incoming links and outgoing links are distributed similarly. On the web, however, incoming links are much more concentrated

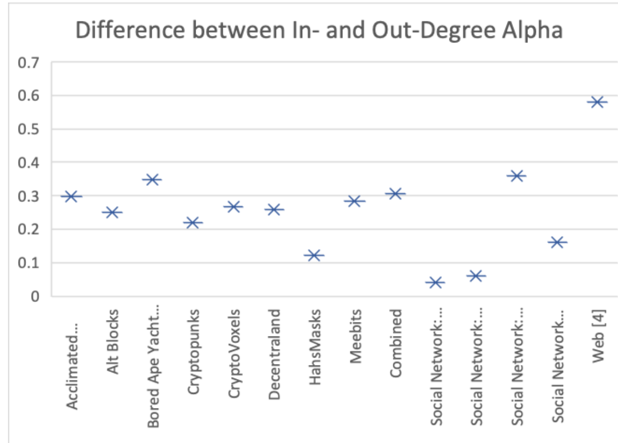

Fig. 6. Difference between In-Degree and Out-Degree Alpha

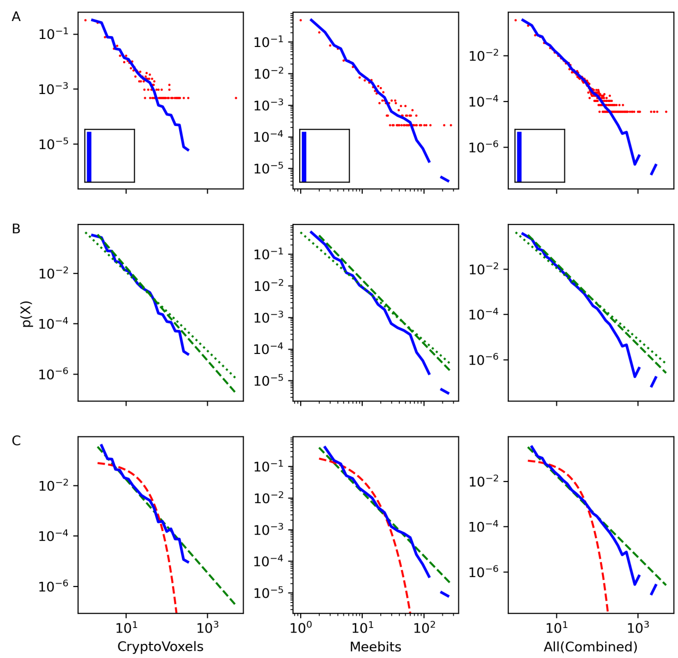

Fig. 7. PDF on log-log axes. A: Blue Line - logarithmically spaced bins, red dots - linear spaced bin; B: Green Dotted Line - power-law fit start at $X_{min} = 1$, Green Dashed Line - power-law fit start from optimal X_{min}; C: Red Dash Line - exponential fitted line

on a small number of high-degree nodes than outgoing links. It implies that there are differences between the populations of active (high out-degree) and popular websites (high in degree) [5]. Studies on how in- and out-degrees are distributed in web networks have frequently contributed to discovering more effective strategies for finding relevant information online [5]. For the NFT network, this indicates the accumulator/NFT collectors are substantially concentrated on a few people, akin to the web influencers.

D. Path Lengths and Diameter

For each NFT network graph and combined network and the NFT combined network graph, in Table VII, we summarize

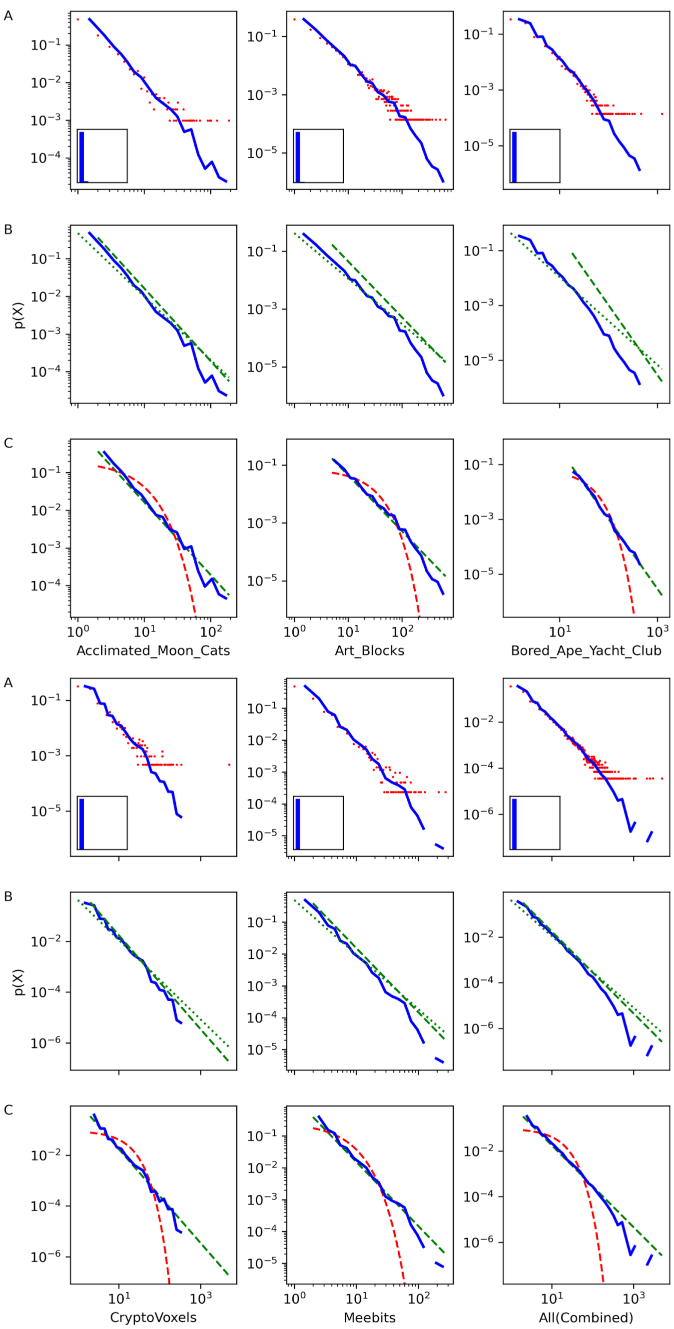

Fig. 8. PDF on log-log axes. Rest of the NFT projects and also the combined NFT network. All suggest good fit to power law.

several metrics for each NFT network graph, including average path length, radius, and diameter.

- **Average Shortest Path Length**: The average of the shortest path lengths for all possible node pairs. Gives a measure of 'tightness' of the Graph and can be used to understand how quickly/easily something flows in this Network [6]
- **Eccentricity**: Eccentricity of a node V is the maximal shortest path distance between V and any other node. [6]

TABLE IV
POWER-LAW COEFFICIENT ESTIMATES (α), AND CORRESPONDING KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT MEASURE (KS) FOR NFT COLLECTIONS.

Collections	**Indegree**		**Outdegree**	
	α	***KS***	α	***KS***
Acclimated Moon Cats	2.313	0.033	2.014	0.033
Art Blocks	1.655	0.046	1.904	0.046
Bored Ape Yacht Club	2.488	0.023	2.139	0.023
Cryptopunks	2.045	0.028	1.825	0.028
CryptoVoxels	2.253	0.029	1.986	0.029
Decentraland	2.318	0.019	2.058	0.019
HashMasks	2.216	0.018	2.095	0.018
Meebits	2.353	0.028	2.07	0.028
Combined	2.167	0.023	1.862	0.023

TABLE V
POWER-LAW COEFFICIENT ESTIMATES (α), AND CORRESPONDING KOLMOGOROV-SMIRNOV GOODNESS-OF-FIT MEASURE (KS) OF WEB, SOCIAL NETWORK: FLICKR, LIVEJOURNAL, AND YOUTUBE.

Networks	**Indegree**		**Outdegree**	
	α	***KS***	α	***KS***
Social Network: Flickr [3]	1.78	0.0278	1.74	0.0575
Social Network: LiveJournal [3]	1.65	0.1037	1.59	0.0783
Social Network: YouTube [3]	1.99	0.0094	1.63	0.1314
Social Network (Combined) [3]	1.81	0.047	1.65	0.089
Web [4]	2.09	-	2.67	-

- **Radius**: the minimum eccentricity across all vertices. [6]
- **Diameter**:the maximum eccentricity across all vertices. [6]

Note that the diameter and average path length closely resemble those found in social network graphs. Both individual and combined transaction graphs demonstrate this. The average path lengths and diameters for all NFT network graph and four social networks are short while those of Web network are much larger. The high degree of reciprocity (mutual links) found in NFT transactions and social networks may be the cause of this feature.

TABLE VI
AVERAGE PATH LENGTH, RADIUS, AND DIAMETER OF THE NFT TRANSCATION NETWORKS.

Network	***Avg. Path Length***	***Diameter***	***Radius***
Acclimated Moon Cats	4.345	12	6
Art Blocks	4.03	12	6
Bored Ape Yacht Club	4.176	12	7
Cryptopunks	3.941	12	6
CryptoVoxels	3.238	9	5
Decentraland	4.104	13	7
HashMasks	4.21	26	13
Meebits	4.885	13	7
Combined	4.461	29	15

To have a better view of the NFT network graph, we also plotted them out using *Pyvis* package. Figure 9 (Left) shows the entire transaction network (mint, buy/sell, and transfer) of the Hash Mask. Noticed that nodes at the outside are disconnected from the center, which indicates the independent

TABLE VII
AVERAGE PATH LENGTH, RADIUS, AND DIAMETER OF THE SOCIAL NETWORKS (FLICKR, LIVEJOURNAL, YOUTUBE) AND WEB NETWORK.

Collections	***Avg. Path Length***	***Diameter***	***Radius***
Flickr [3]	5.67	13	27
LiveJournal [3]	5.88	12	20
YouTube [3]	5.10	13	21
SocialNetwork [3]	5.55	12.7	22.67
Web [4]	16.12	475	905

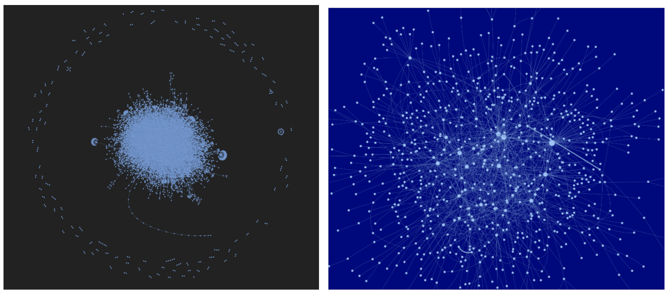

Fig. 9. Left: Hash Mask Entire Transaction Network. Right: Acclimated Moon Cats Largest connected graph

transactions between a few people. These wallets do not contribute much to the collection community.

Figure 10 displays the buy/sell network of CryptoVoxels and Decentraland respectively. By comparing the two graphs, we found that while there are few Decentraland distinct hubs at the center, only one obvious large hub for CryptoVoxels. This suggests that the person who owns that wallet plays a significant role in the CryptoVoxel community.

E. Connectivity and transitivity

The assortativity coefficient represents how nodes are connected to a known property, which is called the degree of homophily in the graph [5]. The value is in the interval: $[-1, 1]$. If assortativity coefficients are in the interval: $[0.6, 1]$, $[0.2, 0.6)$, $[-0.2, 0.2)$, $[-0.6, -0.2)$, $[-1, -0.6)$ respectively, then a graph corresponds to strongly assortative, weakly assortative, neutral, weakly disassortative and strongly disassortative [5]. The reciprocity coefficient evaluates the ratio of mutual connections on a direct graph. That is: if the probability of a given edge $e(v_i, v_j)$ can be estimated, $e(v_j, v_i)$ also exists. In addition, if there is a connection $e(v_i, v_j)$ or $e(v_j, v_i)$ and a connection $e(v_j, v_k)$ or $e(v_k, v_j)$, which is the likelihood of there being a connection $e(v_i, v_k)$ or $e(v_k, v_i)$. It is also called the clustering coefficient.

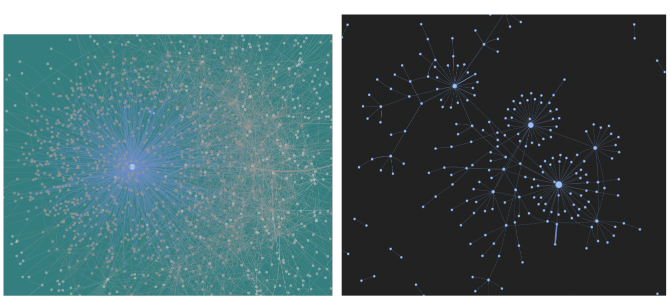

Fig. 10. Left: Buy/Sell Network of CryptoVoxels. Right:Buy/Sell Network of Decentraland

Table 8 shows the assortativity, reciprocity, and clustering of the 8 kinds of NFT. First, the results of assortativity coefficients illustrate that most categories of NFT are neutral associative, while "Acclimated Moon Cats" is strongly associative and "Art Blocks" as well as "Cryptopunks" are weakly associative. Second, the reciprocity coefficients tend to be 0, which illustrates that Buy/Sell transactions are nearly in a single direction. That means exchanges between wallets are rare. Third, the clustering coefficients of the 8 categories also tend to be 0, which implies that collaboration between wallets is weak.

TABLE VIII
THE RESULTS OF CLASSIFICATION ACCORDING TO CONNECTIVITY AND CLUSTERING PROPERTIES.

Collections	***Assortativity***	***Reciprocity***	***Clustering***
Acclimated Moon Cats	0.093647	0.024180	0.014447
Art Blocks	0.041144	0.036908	0.027899
Bored Ape Yacht Club	0.001846	0.054051	0.002710
Cryptopunks	0.163124	0.039362	0.015473
CryptoVoxels	0.106759	0.051368	0.002820
Decentraland	0.206823	0.055868	0.011919
HashMasks	0.234507	0.044048	0.008517
Meebits	0.006323	0.038271	0.005015
TOTAL	0.08506	0.04802	0.008186

F. Power law so what

We discovered through the aforementioned research, in particular the power low, that little changes in a complex system can result in abrupt and massive changes because such small changes cascade down the connected elements of the system, producing a large change. There is also the concept of diminishing returns, which describes the point at which increasing input results in decreasing output. So we propose several suggestions:

- **Hold your winners NFT**: According to the power law, winners perform much better than losers, and everything else is practically useless. Only a small percentage of all projects will ever be considered winners. As a result, if you identify a successful idea, you must stick with it and let it carry you to the top. Due to the power law, the apex is frequently still quite far off and the potential output is still enormous.
- **Network Effects and Positive Feedback Loops.**
- **Hold more NFTs**: Have a wide variety of NFTs in your portfolio to ensure that some of them will satisfy The Power Law. The earnings from your wins will exceed the losses from your losers.

V. CONCLUSION

Focusing on NFT datasets on the Ethereum blockchain, this study has analyzed how the network of transactions performs and distributes. From the graph, it is concluded that the NFT market is decentralized and the degree distribution of many complex networks follows the rule of power law. What's more, there are great similarities between NFT networks and other common networks, especially social networks. In all, this thesis sets out to provide a summary of NFT transaction ecosystem to fill the research gap on related fields.

This study has several limitations, including the scope of applications and the interaction between different NFT products. Furthermore, more researches about NFT network can be done by analogizing the structure and properties of social networks to the structure of complex NFT networks.

ACKNOWLEDGEMENTS

In this work, LIU Zhiruo crawled data and generated datasets, then preprocessed the datasets, then help do one part of graph analysis. ZHANG Keyi was responsible for the methodology part of market analysis. YE Yi Lin did the graph analysis, analysis of network structure. ZHANG Lu was responsible for the characteristics of NFT market. DONG Wenyu drew the pictures of market analysis and completed introduction and conclusion.

REFERENCES

[1] F. Vogelsteller and V. Buterin. EIP-20: ERC-20 Token Standard.[Online]. Available: https://eips.ethereum.org/EIPS/eip-20

[2] F. Victor and B. K. L¨uders, "Measuring ethereum-based erc20 token networks," in Financial Cryptography and Data Security, I. Goldberg and T. Moore, Eds. Cham: Springer International Publishing, 2019.

[3] Ravasz, Erzsébet, and Albert-László Barabási. "Hierarchical Organization in Complex Networks." Physical Review E, vol. 67, no. 2, 14 Feb. 2003, regan.med.harvard.edu/pdf/publications/Hier-Lect-Notes-Phys.pdf, 10.1103/physreve.67.026112. Accessed 26 Apr. 2021

[4] J.Alstott, E.Bullmore, D.Plenz, "powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." PLoS ONE 9(1): e85777. doi:10.1371/ journal.pone.0085777. January 29, 2014

[5] A.Mislove, M.Marcon, K.P.Gummadi, P.Druschel, and B.Bhattacharjee, "Measurement and analysis of online social networks," in Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 29–42.

[6] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph Structure in the Web: Experiments and Models. In Proceedings of the 9th International World Wide Web Conference (WWW'00), Amsterdam, May 2000.

[7] Gruzd, A., Haythornthwaite, C. (2013). Enabling community through social media. Journal of medical Internet research, 15(10), e248. https://doi.org/10.2196/jmir.2796

[8] M. Grandjean. "Introduction to Social Network Analysis: Basics and Historical Specificities." Presented at the HNR+ResHist 2021, Zenodo. https://doi.org/10.5281/zenodo.5083036. 2021.

[9] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[10] bitsCrunch, NFTs: Who's buying them and why? – All questions answered! https://bitscrunch.com/blogs/know-whos-buying-nfts, August 12, 2022

[11] Statistics How To, Zeta Distribution (Zipf Distribution), https://www.statisticshowto.com/zeta-distribution-zipf/