Alphacore - Analysis of Token Liquidity Distribution

Hun Liang
hunliang@vt.edu
Virginia Polytechnic Institute and State University
Blacksburg, Virginia, USA

Baiyi Zhang
baiyi@vt.com
Virginia Polytechnic Institute and State University
Blacksburg, Virginia, USA

James de Chutkowski
jwdchutkowski@vt.edu
Virginia Polytechnic Institute and State University
Blacksburg, Virginia, USA

Abstract

This project proposes an analysis of token liquidity distribution and transaction patterns between centralized (CEX) and decentralized (DEX) exchange addresses within the Ethereum ecosystem that comprises numerous standard ERC20 compliant tokens. Using a data set of more than 6 million addresses and 18 million transaction logs of tokens traded between 2018 and 2020, we examine how token liquidity flows between different exchange types and identify potential market vulnerabilities. Recent studies have confirmed that cryptocurrency liquidity is increasingly concentrated in fewer exchanges and addresses forming systemic risks. Our analysis will map the transaction hierarchies, revealing distinct centralization patterns that align with liquidity concentration documented in the financial literature, and we will quantify how the concentration in liquidity creates market risks through two folds: increased vulnerability to price manipulation and creation of regulatory blind-spots between types of cryptocurrency exchange.

Keywords

Cryptocurrency Exchange Networks, Liquidity Concentration, Token Flow Analysis, Market Stability Risk, ERC20 Transactions, Network Visualization

ACM Reference Format:
Hun Liang, Baiyi Zhang, and James de Chutkowski. 2025. Alphacore - Analysis of Token Liquidity Distribution. In *Proceedings of Social Media Analysis Final Project (Social Media Analysis).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

1.1 Problem Description

Introduction to Problem and Data. This study addresses the growing concerns of liquidity centralization within the Ethereum ecosystem, particularly focusing on token transactions involving centralized (CEX) and decentralized (DEX) exchanges. As cryptocurrency markets mature, liquidity has shown an increasing concentration in a limited number of entities, posing systemic risks such as price manipulation and regulatory blind spot [3]. The object of this project is to analyze transaction patterns and liquidity distribution to identify market vulnerabilities and understand how liquidity flows across different types of exchanges.

1.2 Data Loading and Preprocessing

Using NetworkX, a directed, weighted graph was constructed where nodes represented wallet addresses and edges corresponded to token transfers. Edge weights reflected the value of the transferred tokens. Node attributes included the assigned exchange type, which facilitated later comparative analysis.

1.3 Liquidity Flow Analysis

Overview. Metrics on the inbound liquidity were computed by aggregating the weighted in-degree for each node. These values were used to compare liquidity distributions across CEX and DEX nodes. A boxplot in Figure 1 was generated to visualize the distribution of inbound liquidity volumes across exchange types. The results indicated that centralized exchanges exhibited slightly higher inbound liquidity compared to decentralized exchanges, though the distributions were relatively similar. Unknown addresses demonstrated a broader range with numerous high-value outliers, reflecting the presence of significant yet unlabeled participants. Outbound liquidity patterns were not assessed in this part of the analysis.

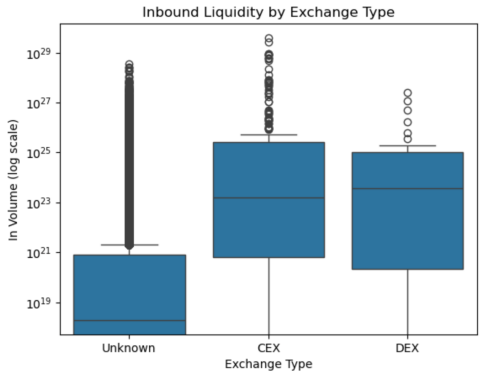

Figure 1: Inbound liquidity (log-scaled volume) by exchange type. The box-and-whisker plot compares Unknown addresses, CEX wallets, and DEX wallets; circles denote outliers above the upper whisker.

2 Network Analysis

2.1 Topological Analysis

Overview. To enable network graph visualization of Alphacore transactions, the original dataset [7] comprising 18 million records

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Social Media Analysis, Blacksburg, VA
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

was subjected to a targeted filtering protocol. Utilizing stream processing, transactions involving addresses classified as CEXs, DEXs, or other predefined known entities were selectively retained. This procedure yielded a computationally tractable edge list, representing approximately 0.5% of the initial data volume, while preserving the categorical CEX/DEX labels associated with the pertinent addresses.

Data Sampling. The analysis dataset was preprocessed by filtering out any entries linked to the zero address (represented as 0x00). To manage computational complexity while preserving network characteristics, a sampling procedure was applied. For each distinct subgraph identified in the data, a 0.5% representative sample was generated. This was achieved using the sample method from the pandas.DataFrame library, specifying `random_state=42` to ensure the sampling process is deterministic and reproducible.

Sample Composition. The underlying transaction data exhibits a predominance of CEX addresses. This skew is maintained within our study sample, comprising 13,043 CEX nodes and 1,766 DEX nodes (7.34x more). Due to the substantially larger node count in the CEX subset, visualizations of the corresponding graph structures will necessarily display greater node density relative to DEX graphs when presented at the same scale.

Average Weighted Degree. The generated directed subgraphs were imported into the Gephi network analysis platform. A key metric calculated for each subgraph was the average weighted degree $\langle k_w \rangle$.This metric provides a measure of the average "strength" of nodes in the network, considering both the number of connections and their weights. It was calculated using the formula: $\langle k_w \rangle = \frac{W}{n}$ In this equation, W denotes the total weight of all edges in the subgraph, where edge weights correspond to the frequency or count of transactions between nodes. n represents the number of nodes (vertices) within the subgraph. The computed average weighted degree $\langle k_w \rangle$ served as a basis for node representation in the subsequent network analysis. Specifically, the visual size of nodes in network diagrams was mapped to their respective subgraph's $\langle k_w \rangle$ value, as shown in the Table 1.

Table 1: Summary statistics for the three exchange-type subnetworks.

Metric	Mixed	CEX	DEX
# of nodes	14 794	13 043	1 766
# of edges	16 746	14 384	2 259
Average Weighted Degree	1.132	1.103	1.279

Fruchterman-Reingold Layout. The spatial arrangement of nodes (representing blockchain addresses) and edges (representing token transactions) in the network graphs was determined using the Fruchterman-Reingold algorithm [2], implemented within Gephi. This force-directed layout simulates physical forces to organize the graph visually. Attractive forces, analogous to springs, are applied along edges, drawing addresses that transact with each other closer together, thus grouping related activity within the visualization. Concurrently, repulsive forces, akin to electrostatic charges, act between all pairs of nodes. This mechanism ensures that nodes maintain a minimum separation distance, preventing visual clutter and overlap, which is particularly important for clarifying structure around high-density areas potentially involving central entities like hub exchanges.

2.1.1 *Modularity Analysis.*

Modularity Maximization Algorithm. Network community structures were determined using the Blondel et al. [1] modularity maximization algorithm for mixed, CEX, and DEX transaction datasets (Figure 1). Node color denotes community classification.

Finding in CEX. Topological analysis reveals shared structural characteristics between the mixed (Fig. 2a) and CEX (Fig. 2b) networks. Both are characterized by a large, predominant community (depicted in blue) centered around a primary hub node, reflecting the significant presence of CEX-associated addresses within the analyzed 'alphacore' dataset. However, the inclusion of bridging interactions within the mixed network influences the community detection outcome. Specifically, the second-largest community in the mixed network constitutes a considerably larger fraction of the total nodes (11.47%) compared to the CEX network (7.85%). This observation suggests that bridging transactions, potentially mediated by DEX platforms, interconnect distinct CEX hubs, thereby integrating them into larger, composite communities within the mixed network representation.

Finding in DEX. Conversely, the DEX network (Fig. 2c) exhibits a distinct topological profile. It lacks a single dominant community, instead displaying several principal communities of more comparable size. This structural difference is corroborated by the Fruchterman-Reingold force-directed layout, which illustrates a relatively uniform spatial distribution of DEX hub nodes, indicative of potentially higher interconnectivity among them (Fig. 2c). In contrast, the CEX hubs (Fig. 2b) demonstrate a more clustered and irregular spatial arrangement, suggesting lower levels of inter-community linkage. Collectively, these findings point towards disparate topological roles fulfilled by CEX and DEX interactions within the transaction ecosystem under investigation.

Modularity Score. Quantitative assessment of the detected partitions was performed using the maximum modularity score (Q). Modularity values approaching the theoretical maximum of 1 signify a network structure with strong community separation. The calculated scores for all three networks (Mixed: Q=0.799, CEX: Q=0.744, DEX: Q=0.756) are substantially high, indicating robust community partitioning in each case. Notably, the Mixed network achieved the highest modularity score. This suggests that while bridging interactions merge some hub-centric structures as observed visually, the resulting overall network partition remains exceptionally well-defined, potentially due to these bridges reinforcing the boundaries of the newly formed, larger communities.

2.1.2 *Inner-Community Analysis.*

Binance Transactions. We performed topological analysis using Gephi on the most central hubs within our transaction network. Specifically, we selected Binance addresses from the labeled dataset and generated a subgraph representing 0.5% sampled transactions

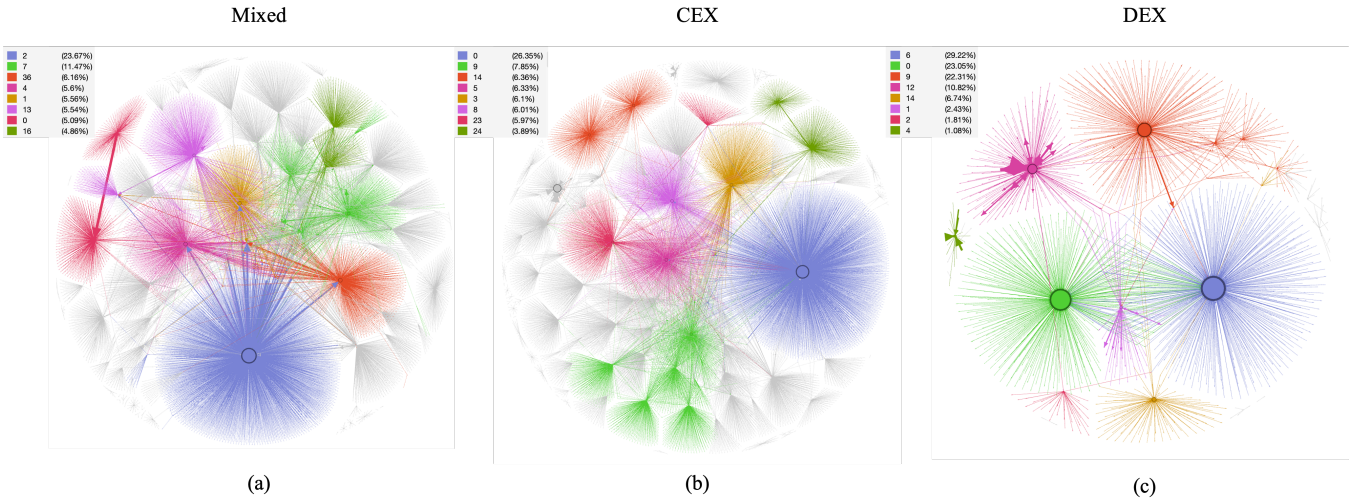

Figure 2: Louvain community layouts for the three subnetworks— (a) Mixed, (b) CEX, and (c) DEX. Node colour denotes modularity class; larger nodes highlight the most-active exchange wallets.

from the alphacore data, encompassing all activities related to the 11 labeled Binance addresses. Utilizing the modularity maximization algorithm, the resulting graph was partitioned into four distinct modules, as depicted in Fig. 4. The central nodes within these modules correspond to Binance addresses [8] labeled as Binance 1, 3, 4, and 2, ranked according to their weighted degrees in Table 2.

Table 2: Weighted connectivity and centrality statistics for key Binance wallets.

Label	WID	WOD	WD	PR	Mod.
BN 1	3 431	1 659	5 090	0.206	0
BN 3	61	1 218	1 279	0.007	1
BN 4	54	1 146	1 200	0.006	2
BN 2	50	1 120	1 170	0.005	3

Symmetry. Our topological analysis reveals a highly symmetrical network structure, with an evident symmetry axis labeled as l in Fig. 3. This symmetry suggests a systematic approach to load balancing within the centralized exchange (CEX) environment, governed predominantly by transaction counts. The strict hierarchical organization of these modules implies limited resilience to external disruptions or turbulence, highlighting potential vulnerabilities within this configuration. Although four prominent communities have emerged around the Binance addresses, with Binance 1 notably dominating the PageRank scores (Table 2), the symmetrical interdependence indicates that functionality across all four hubs is critical. Consequently, despite the strategy of employing multiple addresses, our findings suggest inherent risks remain due to significant dependencies among these interconnected hubs.

2.1.3 *Link Analysis.*

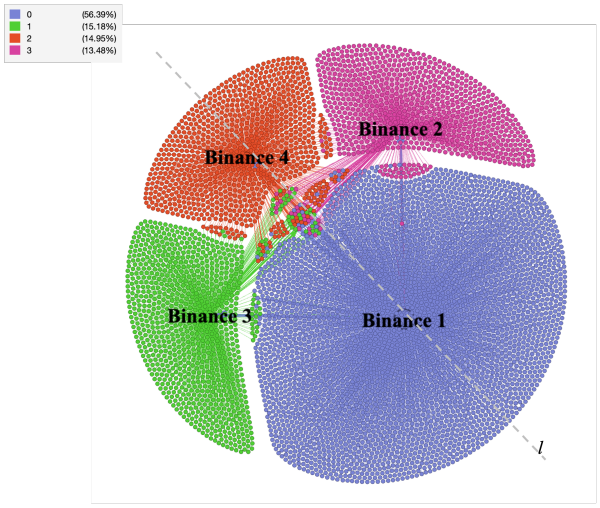

Figure 3: Inner-community structure around the four largest Binance wallets. Each color denotes a modularity class (legend, upper left). Dashed line ℓ highlights the near-symmetrical load-balancing axis connecting the hubs.

PageRank. The PageRank algorithm [6] was applied to quantitatively evaluate the relative importance of nodes within three distinct cryptocurrency transaction network graphs. PageRank scores reflect a node's prominence based on both direct connections and its position within the broader network topology. In line with standard practices in network analysis, we selected a damping factor of 0.85. This parameter represents the probability that a random "surfer" continues following links between nodes, effectively balancing local transactional interactions and the global network structure. To ensure convergence efficiency without compromising the accuracy of results, we set the tolerance threshold at $\epsilon = 0.001$.

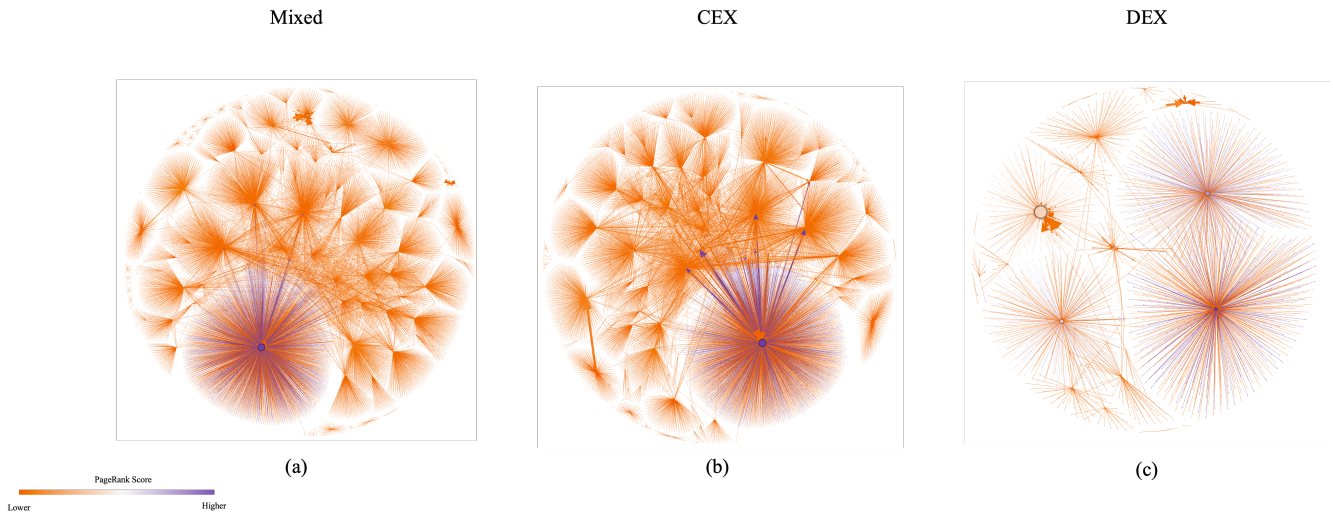

Figure 4: PageRank score distribution in the three subnetworks—(a) Mixed, (b) CEX, and (c) DEX. Node colour follows the scale (orange = lower, purple = higher); larger nodes highlight wallets with the greatest PageRank values.

Centralized Distribution of Influence in CEX. As depicted in Fig. 4, the mixed transaction network (a) and centralized exchange (CEX) network (b) exhibit similar structural characteristics: the majority of nodes have low PageRank scores, indicating limited significance, whereas a few hub nodes dominate with significantly higher scores. Conversely, the decentralized exchange (DEX) network (c) displays a comparatively uniform distribution of PageRank scores across nodes. Notably, in Fig. 4(c), high PageRank values do not directly correlate with node activity levels (represented visually by node size indicating weighted degree, or transaction frequency).

Decentralized Distribution of Influence in DEX. In the CEX network, liquidity and influence are markedly concentrated around a singular, dominant exchange wallet, indicating substantial centralization and potential systemic risks. This centralization implies a vulnerability, as disruption of the dominant node could significantly impact network functionality. In contrast, the DEX network demonstrates a more decentralized distribution of influence across multiple addresses, suggesting shared control over token flows and reduced susceptibility to risks associated with single-point failures.

2.2 Exchange Type-Centric Subgraph Analysis

2.2.1 *Topological Exploration on Binance Nodes.*

Directed Subgraph Extraction. To further investigate localized liquidity concentration, we constructed a focused transaction subgraph Fig. 5 centered around two of the most active Binance exchange addresses in our labeled dataset: 0x3f5ce5... and 0xd55123.... These addresses are commonly referenced as central clearing points within Ethereum token flows. By extracting each node's direct predecessors and successors from the full transaction graph G, a pruned subgraph H was created, capturing the immediate transaction neighborhood of the selected Binance wallets. This approach isolates high-relevance interactions and eliminates noise from unrelated token activity.

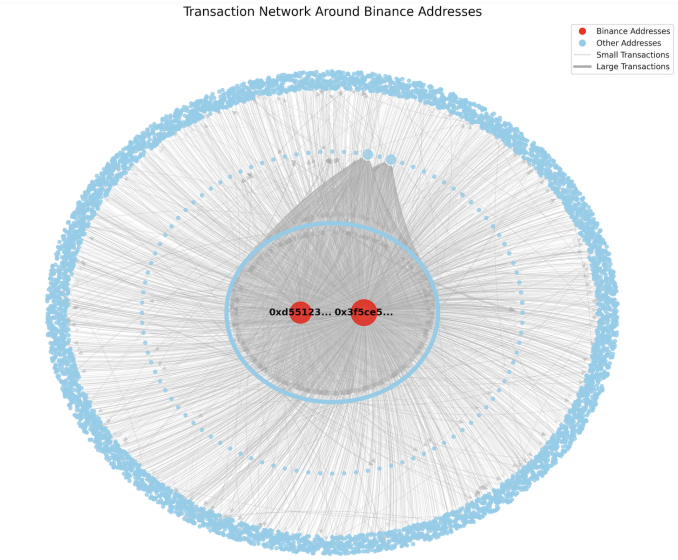

Figure 5: Directed subgraph centered on two primary Binance exchange addresses. Node color distinguishes Binance wallets (red) from all other addresses (blue)- node size proportional to weighted out-degree. Edges represent token transfers, filtered to retain the top 50% by frequency. Curved arrows indicate direction and relative transaction volume.

Node-Edge Encoding. To encode structural prominence, nodes were visually scaled by their weighted out-degree. Binance-controlled addresses were highlighted in red, while all other addresses were marked in sky blue. Edge weights—based on transaction frequency—were linearly rescaled to determine line thickness, with darker gray tones

indicating high-frequency (large) transactions and lighter tones denoting lower-activity ones. To emphasize the most informative interactions, only transactions in the top 50th percentile by weight were retained for visualization.

Hierarchical Layout and Spatial Logic. We employed a three-tiered radial layout to show each node's relative proximity and transactional relevance to the Binance hubs. The central ring contains the seed Binance nodes, surrounded by a secondary ring of nodes with strong bilateral connectivity to both hubs. A tertiary outer ring hosts less-connected addresses that form the long tail of the transactional distribution. Random perturbation/variation and angular spacing techniques were applied to minimize overlap and improve readability. This structured spatial logic reveals a distinct hub-and-spoke configuration, a typical sign of centralized exchange-mediated flow.

Edge Filtering and Visual Density Control. To address the challenge of edge overplotting in dense subgraphs, transaction edges were filtered using a percentile-based threshold, retaining only 4,444 of the original 8,665 edges (approximately 51%). The visual impact of retained edges was modulated through curvature and arrows, aiding the interpretation of token flow directionality. This filtering step reduced noise while preserving essential topological patterns.

Observations. The resulting subgraph exhibits strong signs of transactional centralization. Both Binance addresses operate as dominant hubs, exhibiting high out-degree and mediating thousands of direct token transfers. A small subset of neighboring nodes maintain moderate connectivity—suggesting they may function as relayers or smart contract aggregators[9]—whereas the outermost nodes appear as single-interaction participants. This distribution reflects asymmetric flow behavior consistent with previous studies of centralized exchanges, where liquidity is funneled through a small set of strategically placed wallets.

2.2.2 *Liquidity Rank Distribution Analysis.*

To further characterize the centralization of liquidity within the Binance-centric transaction subgraph, we examined the rank-order distribution of both incoming and outgoing liquidity across labeled addresses. By ranking entities based on liquidity magnitude and applying a base-10 log-log transformation to both rank and value axes, we visualized the tail behavior of these distributions in Fig. 6.

Observations. The resulting log-log plots (Fig. 6) strongly suggest a power-law distribution in both incoming and outgoing liquidity across addresses, particularly among CEX-linked entities. Such power-law behavior implies liquidity centralization, where a small number of actors control a disproportionately large share of token flows. This aligns with systemic risk indicators commonly found in complex financial systems: central hubs not only process the majority of volume but also become critical points of failure [4]. These findings reinforce earlier insights from our modularity and centrality analyses, emphasizing that Binance's transaction structure reflects a highly centralized architecture, consistent with CEX behavior.

2.2.3 *Cross-Type Transaction Flow Analysis.*

To learn how liquidity traverses between distinct categories of exchange entities, we computed the aggregate transaction volume from and to each exchange type using address-level labels. This analysis quantifies the directional intensity of token transfers among centralized, decentralized exchanges, and other unclassified or miscellaneous addresses.

Data preparation & Processing. The dataset was first normalized by converting all Ethereum addresses to lowercase and deduplicating any repeated address-type mappings. We then merged the from-address and to-address fields in the transaction log with the known entity classifications from the liquidity dataframe, defaulting to 'OTHER' where a match could not be found. After merging, the missing type fields were filled accordingly. The resulting transaction flow matrix in Fig. 7 was constructed by grouping the data by (from-type, to-type) pairs and computing the frequency of each transition. These frequencies were reshaped into a pivot table and visualized as a heatmap to emphasize asymmetries and dominant pathways of liquidity transfer

Observations. The heatmap reveals a highly asymmetric flow dynamic. A dominant share of CEX-initiated transactions (101,485) are directed toward Other addresses, with relatively negligible transfer activity toward DEXs. In contrast, DEX addresses only engage in outbound transfers to Other types, with no observed transactions targeting CEXs or other DEXs—likely reflecting operational isolation or limitations in inter-DEX routing within the sampled subgraph.

Strikingly, the Other-Other cell exhibits the highest interaction count (655,659), indicating that a large portion of network activity occurs between addresses outside of recognized exchange categories. This could encompass wallet-to-wallet trading, bridge protocols, or automated contract interactions. Additionally, the Other-CEX and Other-DEX flows (84,614 and 80,672 respectively) show that Other addresses act as key feeders into both exchange types, particularly for liquidity provisioning or token off-ramping [5].

3 Conclusion

Problem. The rapid growth of the Ethereum ecosystem has been accompanied by an equally rapid concentration of token liquidity in a handful of wallets—particularly those controlled by centralized exchanges. Such concentration amplifies systemic risk: a single point of failure or a coordinated manipulation attempt can ripple outward, distorting prices, draining liquidity from smaller venues, and obscuring flows from regulators. Understanding how, where, and why this concentration emerges is therefore a prerequisite for any robust risk-mitigation policy or market-design intervention.

Experiment. To illuminate these dynamics, we assembled what is—to the best of our knowledge—the largest openly analysed ERC-20 transaction graph to date (18M transfers among 6M addresses, 2018–2020). After label-enriching the graph with 13K CEX and 1.8K DEX entities, we applied a reproducible streaming pipeline that (i) prunes noise, (ii) samples representative sub-graphs, and (iii) computes topology-aware metrics in both Python/NetworkX and Gephi. Key experiments included Louvain modularity maximisation, PageRank centrality, weighted-degree scaling, and symmetry analysis on Binance-centric sub-communities. Each step was

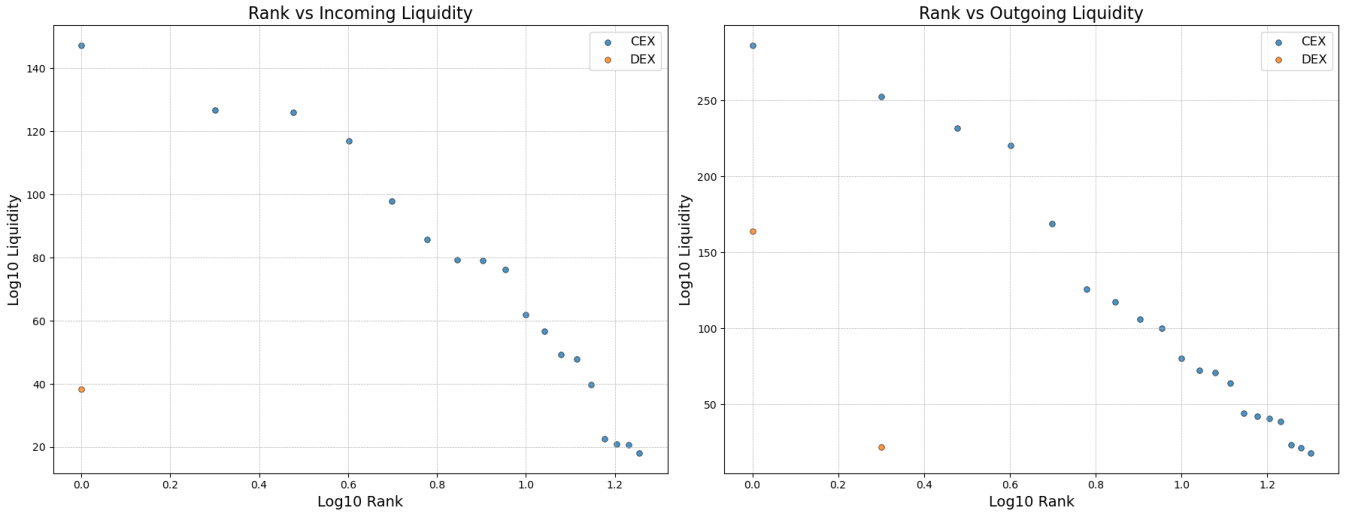

Figure 6: Log-log rank distribution of liquidity volumes within the Binance-centric subgraph. (Left) Incoming liquidity ranked by address. (Right) Outgoing liquidity. The approximately linear trend on the log-log scale suggests power-law behavior, particularly among CEX-associated addresses

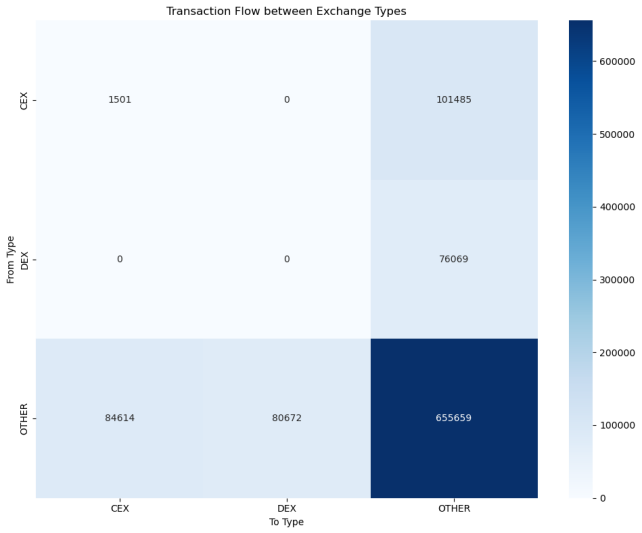

Figure 7: Heatmap showing the frequency of token transfers between address types: CEX (Centralized Exchanges), DEX (Decentralized Exchanges), and OTHER. Values represent transaction counts aggregated by source and destination type.

benchmarked against ground-truth label distributions to validate coverage and minimise sampling bias.

Findings. Across all analyses, CEX wallets exhibit textbook hub-and-spoke behavior: >80% of observed liquidity passes through one dominant address in each cluster, yielding power-law tails and modularity scores above 0.74. DEX networks, by contrast, show flatter PageRank spectra and multiple similarly sized communities, signalling a healthier distribution of influence. Crucially, even when CEXs fragment activity across several labelled wallets (e.g., the four main Binance hubs), strict bilateral symmetry keeps the system vulnerable to hub failures. These results quantify long-suspected risks: today's token markets remain only as resilient as their largest CEX, while DEX infrastructure—despite lower absolute volume—offers a blueprint for decentralised liquidity that is both more diffused and, potentially, more robust.

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[2] Thomas MJ Fruchterman and Edward M Reingold. 1991. Graph drawing by force-directed placement. *Software: Practice and experience* 21, 11 (1991), 1129–1164.

[3] Sascha Hägele. 2024. Centralized exchanges vs. decentralized exchanges in cryptocurrency markets: A systematic literature review. *Electronic Markets* 34 (12 2024), 1–21. doi:10.1007/s12525-024-00714-2

[4] Matthew O. Jackson and Agathe Pernoud. 2020. Systemic Risk in Financial Networks: A Survey. arXiv:2012.12702 [q-fin.RM] https://arxiv.org/abs/2012.12702

[5] Alfred Lehar, Christine Parlour, and Marius Zoican. 2024. Fragmentation and optimal liquidity supply on decentralized exchanges. arXiv:2307.13772 [q-fin.TR] https://arxiv.org/abs/2307.13772

[6] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford infolab.

[7] Kiarash Shamsi, Yulia R. Gel, Murat Kantarcioglu, and Cuneyt G. Akcora. 2022. Chartalist: Labeled Graph Datasets for UTXO and Account-based Blockchains. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, November 29-December 1, 2022, New Orleans, LA, USA*. 1–14. https://openreview.net/pdf?id=10iA3OowAV3

[8] Kiarash Shamsi, Friedhelm Victor, Murat Kantarcioglu, Yulia Gel, and Cuneyt G Akcora. 2022. Chartalist: Labeled Graph Datasets for UTXO and Account-based Blockchains. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 34926–34939. https://proceedings.neurips.cc/paper_files/paper/2022/file/e245189a86310b6667ac633dbb922d50-Paper-Datasets_and_Benchmarks.pdf

[9] Takaaki Yanagihara and Akihiro Fujihara. 2025. Relayer Aggregation Using Chainless Multi-Layer Consensus. *IEEE Access* (04 2025). doi:10.1109/ACCESS.2025.3561243