Pseudonymity at Risk: Linkage Attacks on Blockchain Users with Off-Chain Cues

Alexander Smirnov
TU Berlin
Berlin, Germany
alex.smirnov@web.de

Stefan Schmid
TU Berlin and Weizenbaum Institute
Berlin, Germany
stefan.schmid@tu-berlin.de

Friedhelm Victor
TRM Labs
San Francisco, USA
friedhelm@trmlabs.com

Abstract

Public blockchain pseudonymity is vulnerable to "off-chain cues": external information from sources such as social media or data leaks that can be linked to on-chain activity, enabling user de-anonymization. This paper provides the first comprehensive framework for systematically investigating this threat. We introduce a taxonomy of off-chain cues and propose methodologies to match such cues to blockchain data. Our empirical evaluation on the Ethereum blockchain quantifies the de-anonymization potential of both individual cues and their combinations, as well as real-world case studies such as the Celsius data leak. The results show that cue effectiveness is largely determined by uniqueness and exactness (e.g., specific transaction quantities, interactions with rare tokens), and that combining even a small number of moderately specific cues can significantly reduce anonymity sets. For users revealing multiple such cues, this can lead to the identification of a single on-chain address. In our Celsius simulation, knowledge of only the transaction day, asset, and exact quantity was sufficient to potentially identify over 85% of depositors. Our findings demonstrate the significant privacy risks posed by off-chain cues in blockchain linkage attacks to users disclosing them. Our work underscores the need for heightened user awareness and the development of effective countermeasures.

Keywords

Blockchain, De-anonymization, Linkage Attack

1 Introduction

In October 2022, during the bankruptcy case of the cryptocurrency platform Celsius Network, documents were released that included spreadsheets detailing customer transaction histories within the platform [8]. The records contained customer names connected to details of Celsius-internal transactions and external transactions (deposits/withdrawals to external wallets on-chain). Although blockchain user addresses were redacted, activity timestamps were only truncated to dates and exact amounts and assets were disclosed. Users on Reddit were quick to notice that this enables de-anonymization and highlighted the risk that "criminals and scammers can find out which people still have crypto on the balances connected to their names[...]"[1]. The disclosure represents a large-scale instance in which external information likely allowed linking identities to specific activities on the blockchain (on-chain).

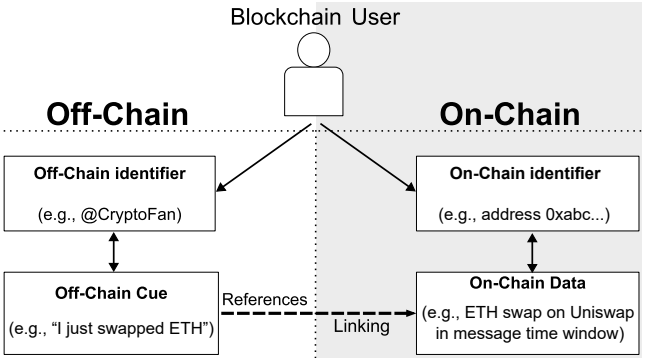


Figure 1: The dashed arrow highlights the paper's core idea: off-chain cues (e.g., social media posts or leaked records) can be linked to on-chain activity, enabling connections between blockchain addresses and real-world identities.

The case of Celsius Network highlights a fundamental privacy issue of most public blockchains: The pseudonymity of users can be compromised when external information is correlated with public on-chain data. Once an address is linked to an identity, the individual becomes vulnerable to a range of targeted attacks, including phishing schemes [5, 11, 20], extortion [16], social engineering [22], and potentially even physical attacks[15].

However, the release of these documents is only one well-known instance; the issue extends to social media, where users sometimes share information about their on-chain activities in off-chain contexts, creating what we term *off-chain cues*. These cues, when correlated with the publicly available on-chain data, can significantly reduce or compromise the pseudonymity, potentially leading to user identification. Our threat model assumes access to off-chain cues. Users who disclose none are not subject to this attack.

Existing studies have already demonstrated the feasibility of de-anonymization by linking specific types of external information, such as social media posts containing addresses, known counterparty interactions, transaction details and IP addresses, to on-chain activity [1, 3, 6, 7, 17, 19, 23]. However, a comprehensive framework that systematically categorizes off-chain cues linkable to on-chain identifiers and quantifies their individual and combined de-anonymization potential had yet to be developed. Unlike prior work using explicit identifiers, our framework studies indirect off-chain cues and their combined deanonymization impact.

[1] https://www.reddit.com/r/btc/comments/xxtfuv/comment/irdswzm/

Proceedings on Privacy Enhancing Technologies 2026(3), 787–800

https://doi.org/10.56553/popets-2026-0106

This paper systematically investigates how off-chain cues can compromise the pseudonymity of blockchain users. As visualized in Figure 1, the core idea is that a user's off-chain data can be linked to their on-chain activity, enabling de-anonymization.

To this end, we make the following contributions:

(1) We introduce a taxonomy of off-chain cues in Section 3, categorized into transaction- and account-based cues, allowing for systematic analysis.
(2) In Section 5, we empirically evaluate the de-anonymization potential of individual cues and various combinations of cues, identifying which cues are most effective at reducing the anonymity sets. Our evaluation shows that cues with high uniqueness (e.g., rare tokens, non-round amounts) and high exactness (e.g., exact timestamps, full address strings) can very effectively compromise user pseudonymity. We demonstrate that combining only a few moderately specific cues can lead to a significant reduction in anonymity sets, sometimes narrowing them down to fewer than a hundred candidates or even enabling unique identification.
(3) We apply our methodologies to synthetic scenarios like the Celsius data leak in Section 6, showing the practical feasibility and implications of cue-based de-anonymization. Our simulation of the Celsius data leak reveals that with knowledge of the transaction day, asset, and exact quantity, over 85% of depositors could potentially be identified.
(4) We develop a LLM-based methodology to extract off-chain cues from unstructured text data like social media posts in Section 7. This extractor enables a scalable analysis of off-chain cue prevalence on X (formerly Twitter), where we find that approximately 24% of crypto-related tweets contain identifiable cues, with each tweet having an average of 3.63 cues, and asset-related cues are most common.
(5) We develop an LLM-based processing pipeline for automated off-chain cue extraction from text and release the analysis code and data processing pipeline to support reproducibility at https://github.com/LatentSpaceExplorer/de-anonymizing-blockchain-users-through-off-chain-data-linkage.

Finally, we discuss the implications of our findings in Section 8, including how off-chain information disclosure affects users under our threat model.

2 Background and Threat Model

A key characteristic of public blockchains like Bitcoin and Ethereum is that they are pseudonymous, rather than fully anonymous [24]. Users interact with the network through public addresses. In account-based chains, these are known as Externally Owned Accounts (EOAs), which are not directly linked to a user's personal identity. However, the public and immutable nature of these ledgers means that anyone with a block explorer can review transaction histories, analyze address interactions, or trace the flow of cryptoassets [4]. The entire privacy model, therefore, relies on maintaining the separation between an EOA and its operator's real-world identity. If this separation breaks and a link is established, the user's entire on-chain history becomes de-anonymized.

Threat Model. The goal of an adversary performing a linking attack is to deanonymize a target by associating a known off-chain identity (e.g., a social media profile, or a name in a public data leak) with its corresponding on-chain EOA. The successful linkage would expose the target's entire on-chain financial history.

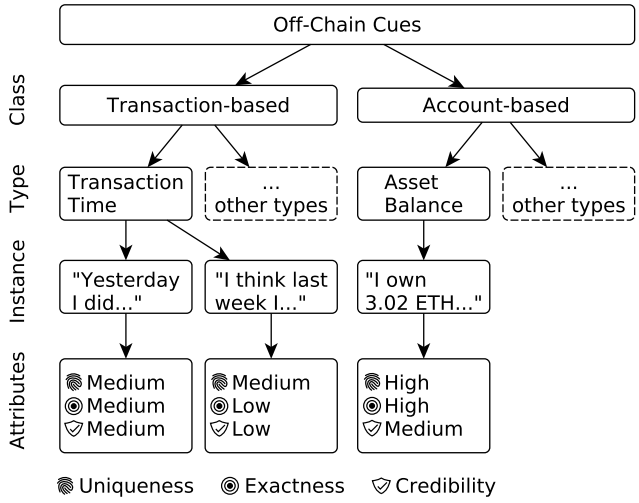


Figure 2: Taxonomy of off-chain cues, categorized into transaction-based and account-based classes. Their de-anonymization potential depends on uniqueness, exactness, and credibility of the information provided.

We assume a passive adversary whose capabilities are limited to the observation and analysis of publicly available information. This adversary is assumed to possess two sets of data:

(1) **On-Chain Data:** The adversary has full access to the public blockchain ledger, allowing for the analysis of all historical transactions, account states and traces, including some label information on addresses that are owned by exchanges or protocols within Decentralized Finance (DeFi) [2].
(2) **Off-Chain Cues:** The adversary collects external information from public sources, such as social media or data leaks, that represent cues that can be linked to on-chain data.

This attack does not involve any active network interaction, such as sending transactions to a target. It is completely passive and cannot be detected by monitoring on-chain activity. However, the attack can be further exacerbated by an attacker performing intentional social engineering, to lead a suspect to reveal information that can become cues for de-anonymization.

We emphasize that our threat model assumes the adversary has access to off-chain cues: users who do not reveal any cues are not subject to this linkage attack.

3 Taxonomy and Matching Methodologies

To systematically investigate the de-anonymization potential of off-chain cues, we developed a taxonomy to classify cues, matching algorithms to link cues with on-chain data, and an automated cue extraction methodology for unstructured text.

Table 1: Transaction-Based Cue Types

Cue Type	Data Type
`Blockchain Used`	Categorical
`Transaction Hash`	Hash
`Transaction Time`	Temporal
`Asset Transferred`	Categorical
`Quantity Transferred`	Numeric
`Fiat Value Transferred`	Numeric
`Counterparty Address`	Hash
`Protocol Interaction`	Categorical
`Transaction Fees`	Numeric
`Transaction Type`	Categorical

Table 2: Account-Based Cue Types

Cue Type	Data Type
`EOA Address`	Hash
`First Activity Time`	Temporal
`Last Activity Time`	Temporal
`Asset Used/Held`	Categorical
`Asset Balance`	Numeric
`Asset Fiat Value`	Numeric
`EOA Total Fiat Value`	Numeric
`Fiat Value Change`	Numeric
`Service Used`	Categorical
`Activity Pattern`	Categorical

Note. The cue types were compiled from real-world off-chain disclosures (e.g., social media posts and leak documents) and commonly available blockchain fields. This list is not exhaustive; for example, NFT token IDs or collection names can be treated as `Asset Transferred` or `Protocol Interaction` cues.

3.1 A Taxonomy of Off-Chain Cues

An "Off-Chain Cue" is a specific piece of external information that can be used to correlate with on-chain data to link a blockchain user's identity to their EOA. To understand and analyze these cues, we developed a hierarchical taxonomy as visualized in Figure 2.
Classes. At the highest level, cues are divided into two classes:

- **Transaction-Based Cues**, which relate to a specific activity on the blockchain.
- **Account-Based Cues**, which describe persistent characteristics of a user's EOA.

Types. Each class contains multiple cue types. Table 1 and 2 lists the specific cue types we have identified within each class. The data type of a cue is fundamental to its de-anonymization potential as it defines the size and distribution of its value space:

- **Hash:** The value space for hashes is very large (e.g., 2^{160} for addresses on Ethereum), meaning each exact instance is effectively unique. Furthermore, the distribution is almost completely even, as the hashes are usually generated without user influence, except for some vanity address generation tools.
- **Numeric:** Numeric values also have a large range of possible values, but the actual distribution of these values on-chain often clusters (e.g., round numbers, common transfer amounts). The uniqueness of a specific numeric cue instance will depend on how common that particular value is.
- **Temporal:** Temporal values also have a large range of possible values, but the on-chain values cluster around the block-timestamps. The uniqueness of such a cue instance depends on the amount of the overall on-chain activity during the time of the instance. Early blockchain history has fewer transactions per block, making temporal cues from that time period more unique.
- **Categorical:** This data type represents a selection from a finite set of categories. For example, assets (e.g., "ETH", "USDT") or DeFi transaction types (e.g., "Swap", "Lend"). The uniqueness of an instance depends on the popularity or rarity of the specific selection from the set.

Instances. Each piece of information is an *Instance* of a cue *Type* (e.g., for the type `Quantity Transferred`, an instance could be "1.502"). The de-anonymization potential of an instance is determined by its attributes, primarily its *Uniqueness* relative to overall blockchain activity, its *Exactness* (e.g., an exact timestamp vs. "last week"), and its *Credibility*. A cue from a verified data leak would have higher credibility than one from an unverified social post. When using scoring (Section 3.2.2), we can assign higher weight to cues with high credibility. The attributes are interdependent: a highly unique cue, such as a non-round transaction amount, only reduces the anonymity set effectively if it also has high exactness and credibility.

3.2 Matching Methodologies

To link the identified off-chain cues with on-chain data, we developed three matching algorithms: (1) *Set Intersection Matching*, (2) *Score-Based Matching*, and (3) *Proximity-Based Matching*. For the quantitative evaluation in this paper, we use the Set Intersection method due to its computational efficiency and its ability to directly measure the remaining anonymity set. Nevertheless, we briefly describe the other two approaches for completeness.

3.2.1 Set Intersection Matching. The Set Intersection method provides a direct and transparent way to quantify de-anonymization. For each cue i, we identify the set of externally owned accounts (EOAs) whose on-chain activity matches the cue, yielding S_i. For N cues, the final candidate set is given by

$$S = S_1 \cap S_2 \cap \ldots \cap S_N \tag{1}$$

The size of the resulting set, $k = |S|$, measures the remaining anonymity set. A value of $k = 1$ indicates potential identification of a unique EOA, while larger sets imply remaining ambiguity.
Note: This strict intersection assumes each user controls a single address. We discuss the multi-address case in Section 8.

3.2.2 Score-Based Matching. To address the brittleness of strict intersections, we developed a weighted scoring approach. Each cue i is assigned a weight w_i based on attributes such as exactness, uniqueness, and credibility. For each EOA, the total score is computed as

$$\text{Score}(e) = \sum_{i:e \in S_i} w_i \tag{2}$$

When weight w_i models cue credibility, then a cue from a confirmed leak might have $w_i = 1$, whereas a low-confidence cue (e.g. rumor) gets $w_i < 1$.

EOAs are then ranked in descending order of $\text{Score}(e)$, with higher scores indicating stronger evidence of matching. This method relaxes the strictness of set intersection and is robust to occasional incorrect or noisy cues.

3.2.3 Proximity-Based Matching. Finally, to handle inherently imprecise cues (e.g., "around 5 ETH" or "late at night"), we extend the score-based approach by introducing similarity functions. For each sortable cue i, we define a similarity function $\text{Sim}(v_{cue}, v_{on\text{-}chain}) \in [0, 1]$ that evaluates how closely an on-chain value matches the cue. The score of an EOA then becomes:

$$\text{Score}(e) = \sum_i w_i \cdot \text{Sim}(v_{cue}, v_{on\text{-}chain}) \tag{3}$$

3.3 Proximity-Based Matching Illustration

For the systematic evaluation in Section 4, we use set intersection, which is well-suited for large-scale analysis when cues are assumed to be valid and exact. However, real-world off-chain cues can be imprecise (e.g., rounded amounts, approximate times) and may carry varying credibility. Strict intersections may eliminate all candidates or remain overly broad. We now provide an illustrative example in which proximity-based scoring yields a meaningful ranking.

Assume an adversary observes the following off-chain post:

> "Started a year ago and kept stacking LINK… Just added close to 1000 units last week."

From this, the adversary extracts the following cue instances:

(1) `Asset Transferred`: LINK,
(2) `Quantity Transferred`: ≈ 1000 LINK,
(3) `Last Activity Time`: "last week",
(4) `First Activity Time`: "started a year ago".

If interpreted strictly (exact timestamp and exact amount), these cues do not intersect on a single EOA. We therefore apply proximity-based matching. For each EOA e, we identify the closest LINK transfer to the reference time (one week before the post) and compute a similarity score per cue. To keep the example simple and reproducible, we use discretized similarity levels:

- **Asset:** 1 if the EOA transferred LINK in the window, 0 otherwise.
- **Amount:** 1 if within ±10% of 1000, 0.5 if within ±50%, 0 otherwise.
- **Time:** 1 if within ±1 day, 0.5 if within ±7 days, 0 otherwise.
- **First activity:** 1 if within ±3 months of one year, 0.5 if within ±6 months, 0 otherwise.

We assign normalized cue weights reflecting exactness, uniqueness, and credibility:

$$w_{\text{asset}} = 0.4, \quad w_{\text{amount}} = 0.25, \quad w_{\text{time}} = 0.2, \quad w_{\text{age}} = 0.15,$$

and compute

$$\text{Score}(e) = \sum_i w_i \cdot \text{Sim}_i(e).$$

Table 3: Illustrative proximity-based ranking for the LINK example. Similarity values in parentheses follow the discretized rules stated in the text; scores are $\sum_i w_i \text{Sim}_i(e)$.

EOA	LINK	Amount	Time	Age	Score	Rank
e_1	yes (1)	1020 (1)	+1d (1)	11 mo (1)	1.000	1
e_2	yes (1)	1400 (0.5)	0d (1)	12 mo (1)	0.875	3
e_3	yes (1)	980 (1)	+5d (0.5)	15 mo (1)	0.900	2
e_4	yes (1)	600 (0.5)	+2d (0.5)	24 mo (0)	0.625	4
e_5	no (0)	1000 (1)	0d (1)	12 mo (1)	0.600	5

Table 3 shows five illustrative candidate EOAs, the raw deviations from the cues, and the resulting scores.

Candidate e_5 matches the temporal and numeric cues well but never interacted with LINK, resulting in a low score despite otherwise favorable attributes. This example illustrates how proximity-based matching degrades gracefully under uncertainty and allows strong but inexact cues to accumulate evidence, rather than being discarded by strict set intersection.

We emphasize that the discretized similarity levels used in this example are intentionally simplistic and chosen for illustrative clarity. In practice, an adversary could employ continuous similarity functions, such as exponential or Gaussian decay for temporal cues, log-scale distance for numeric amounts, or learned similarity models that reflect observed on-chain distributions. These alternatives preserve the same conceptual structure: per-cue similarity weighted by exactness, uniqueness, and credibility, but allow for smoother scoring and finer discrimination between candidates.

4 Data and Evaluation Setup

To quantify the de-anonymization potential of off-chain cues, we evaluate single cues, cue pairs, and targeted case studies (a simulated version of the Celsius leak, multiple known transaction days, multiple known tokens used). The objective is to measure how strongly different cue types, individually and in combination, reduce uncertainty about a user's Externally Owned Account (EOA).

Our analysis focuses on the Ethereum blockchain, using data obtained from Google's public BigQuery Ethereum dataset.[2] This dataset contains the full history of detailed transaction records of the Ethereum blockchain, including timestamps, source and target EOAs, transferred amounts, gas fees, and related metadata. In addition, we use a table on token transfers, provided in the same dataset, which we use to determine asset transfers like the stablecoin USDT and others. Finally, we make use of 11 address labels of the Celsius network, obtained from Etherscan. All evaluation of on-chain cues was conducted using BigQuery SQL queries, and the analysis is fully reproducible with our accompanying source files.

4.1 Baseline

All analyses are conducted on a static snapshot of the Ethereum blockchain with a cutoff date of December 31, 2024. This fixes the initial anonymity set at

$$N = 216{,}376{,}711, \tag{4}$$

[2] https://cloud.google.com/blockchain-analytics/docs/example-ethereum

the number of ever-active EOAs. The corresponding prior entropy (in bits) is

$$H_0 \;=\; \log_2 N \;\approx\; 27.69. \tag{5}$$

An attacker seeks to locate the target's EOA within this set. After applying cues, the remaining indistinguishable candidates form an anonymity set of size k. Note that we assume each pseudonymous user corresponds to a single EOA. In practice, users may split funds, but that more complex scenario is left for future work.

4.2 Evaluation Metrics

We compute the following metrics:

Anonymity Set Size (k): Number of candidate EOAs remaining after applying a cue (or cue set). Smaller k implies stronger de-anonymization.

Identification Rate (IR): Fraction of trials in which the anonymity set collapses to a unique EOA ($k = 1$).

Information Gain (IG): Let a *cue* C narrow the candidate set to k EOAs (within a given sample). The posterior entropy is $H(C) = \log_2 k$, and the (per-sample) *information gain* is

$$IG(C) \;=\; H_0 - H(C) \;=\; \log_2\Big(\frac{N}{k}\Big). \tag{6}$$

For a pair of cues C_0, C_1, let k_0, k_1 be their marginal candidate counts and $k_\cap$ the size of their intersection (within the same sample). The joint information gain is

$$IG_{\text{joint}}(C_0, C_1) \;=\; \log_2\Big(\frac{N}{k_\cap}\Big). \tag{7}$$

As an *independence baseline*, we add the single-cue gains:

$$IG_{\text{indep}}(C_0, C_1) \;=\; IG(C_0) + IG(C_1) \;=\; \log_2\Big(\frac{N^2}{k_0\, k_1}\Big). \tag{8}$$

We report a *mutual-information-style* deviation as the difference between the independence baseline and the actual joint gain:

$$\mathrm{MI}(C_0, C_1) \;=\; IG_{\text{indep}} - IG_{\text{joint}} \;=\; \log_2\left(\frac{N\, k_\cap}{k_0\, k_1}\right). \tag{9}$$

Here, $MI > 0$ indicates redundancy (more overlap than expected), $MI \approx 0$ indicates near-independence, and $MI < 0$ indicates complementarity/synergy (less overlap than expected).

Independence Baseline. Under independence, the expected joint candidate size is

$$\mathbb{E}\big[k_\cap\big] \;\approx\; \frac{k_0\, k_1}{N}, \tag{10}$$

which implies $\mathrm{MI}(C_0, C_1) \approx 0$ on average.

Rationale. We use *information gain* (IG) to express anonymity reduction in bits, yielding a normalized measure that is comparable across cues and additive under an independence assumption. This property enables aggregation of multiple cues and motivates the independence baseline in Eq. (9). The corresponding *mutual-information-style* quantity (MI) captures deviations from this baseline, distinguishing redundant cues from complementary ones. In contrast, reporting only anonymity set sizes (k) does not reveal how cues interact, as identical reductions in k may arise from qualitatively different overlaps. IG and MI therefore provide a clearer characterization of cue effectiveness.

4.3 Aggregation Across Samples

Let $\mathcal{S}$ denote the set of samples with nonempty joint support,

$$\mathcal{S} \;=\; \{\, s \;:\; k_\cap(s) > 0 \,\}, \qquad |\mathcal{S}| = \texttt{count}. \tag{11}$$

We report the following means over $\mathcal{S}$:

$$\overline{IG}_{\text{joint}}(C_0, C_1) = \frac{1}{|\mathcal{S}|} \sum_{s\in\mathcal{S}} \log_2\Big(\frac{N}{k_\cap(s)}\Big), \tag{12}$$

$$\overline{IG}_{\text{indep}}(C_0, C_1) = \frac{1}{|\mathcal{S}|} \sum_{s\in\mathcal{S}} \Big[\, \log_2\Big(\frac{N}{k_0(s)}\Big) + \log_2\Big(\frac{N}{k_1(s)}\Big)\Big], \tag{13}$$

$$\overline{\mathrm{MI}}(C_0, C_1) = \frac{1}{|\mathcal{S}|} \sum_{s\in\mathcal{S}} \log_2\left(\frac{N\, k_\cap(s)}{k_0(s)\, k_1(s)}\right). \tag{14}$$

By linearity of averaging, $\overline{\mathrm{MI}} = \overline{IG}_{\text{indep}} - \overline{IG}_{\text{joint}}$.

Special Cases. For a cue intersected with itself, $k_\cap = k_0 = k_1$, hence

$$IG_{\text{joint}}(C, C) = IG(C), \qquad \mathrm{MI}(C, C) = IG(C). \tag{15}$$

4.4 Experimental Setups

Our evaluation is designed to assess the de-anonymization risk of off-chain cues from multiple perspectives, including the analysis of individual cues, combinations of cues, realistic case studies, and insight into the prevalence of cues on social media:

- **Individual and Combined Cues.** We measured the Information Gain for a large set of individual cue types (e.g., transaction time, asset used, partial address). We then analyzed pairs of individually weaker cues to quantify their combined effectiveness and mutual information (Section 5).
- **Celsius Data Leak.** We simulated an attack using the Celsius leak data. The analysis started from known Celsius deposit addresses and measured the effectiveness of cues such as transaction day, asset transferred, and exact quantity (Section 6.1).
- **Knowledge of Activity Dates.** We evaluated the impact of knowing specific dates on which a target EOA was active. Such information could stem from leaks or off-chain timelines. To focus on typical users, we excluded high-frequency accounts (EOAs with >250 transactions in 2024), to avoid including noise such as bots and exchanges. The experiment assumes an attacker knows d active dates ($d \in [1, 10]$) (Section 6.2).
- **Knowledge of Token Used.** We studied the effect of knowing which tokens a user has sent at least once. Potential sources include portfolio leaks or community affiliations. To avoid rare assets, only tokens used by more than 50,000 EOAs before 2025 were considered. The experiment assumes an attacker knows t tokens ($t \in [1, 10]$) (Section 6.3).
- **Social Media Cue Prevalence.** To assess real-world prevalence, we analyzed 10,000 crypto-related tweets from X (formerly Twitter) we designed an LLM-based cue extractor to automatically identify and categorize off-chain cues (Section 7).

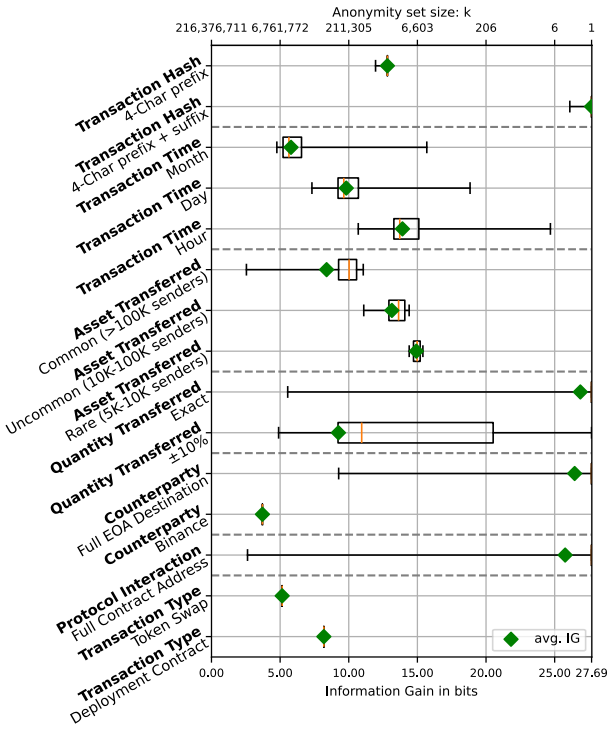


Figure 3: Information gain per *individual transaction cue*. Diamonds show mean IG across samples, gray bars indicate box-plots. Horizontal dashed lines delineate cue groupings. A popular counterparty like Binance yields little IG. Finer-grained timestamps (day/hour) provide markedly more information than coarse ones (year/month). While exact transferred quantity yields high information gain, approximated quantity adds expectedly a lot less. Overall, cues tied to specific addresses (counterparty address, transaction prefix+suffix) and exact quantities transferred) deliver the strongest single-cue narrowing of the candidate set.

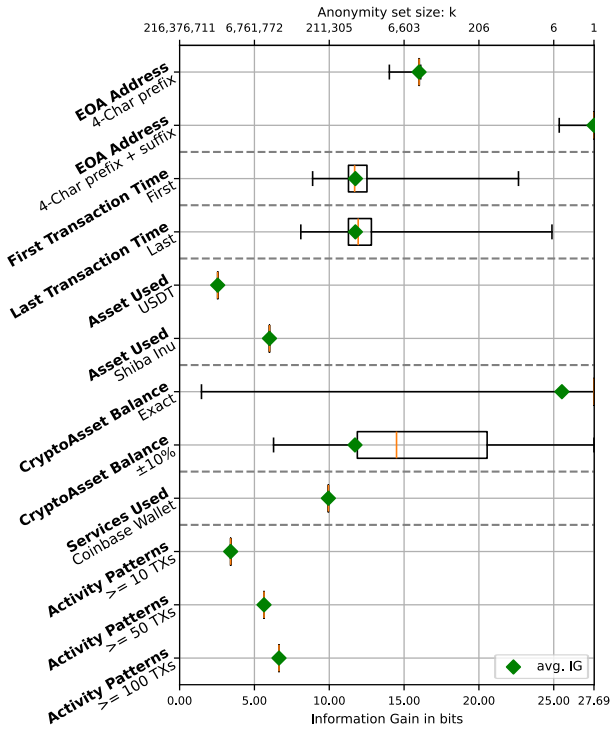


Figure 4: Information gain per *individual account cue*. Diamonds show mean IG across samples, gray bars indicate box-plots. Horizontal dashed lines delineate cue groupings. Highly specific cues like 4-char EOA address prefix/suffix deliver the largest single-cue information gains (on average identifying a single EOA), while temporal cues (First/Last Transaction Day) provide substantial but smaller gains. Asset-usage signals are moderate, and pure activity thresholds (e.g., ≥10/50/100 TXs) contribute only a few bits. Balance-based cues exhibit wide dispersion, reflecting heterogeneity across accounts.

5 Individual and Combined Cue Evaluation

This section presents our empirical evaluation of blockchain de-anonymization, quantifying the potential of different cue types individually and in combination from an attacker's perspective.

5.1 Individual Cues

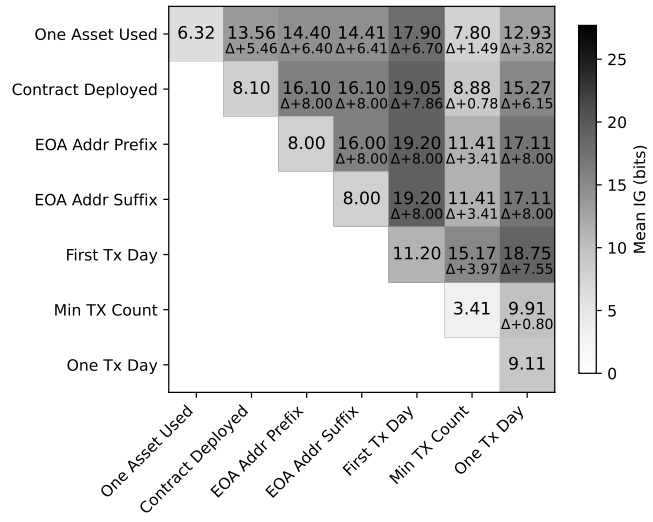
For each cue (e.g., a specific transaction time or token transfer), we identify all EOAs matching that cue on-chain and take k as the size of that candidate set. Figure 3- 4 then show the distribution of IG across many random instances of each cue. Except for a few computationally heavy cases, the full Ethereum dataset was used. Three cue types were instead evaluated on a 1% random sample to reduce runtime complexity:

- EOA 4-Character Prefix + Suffix
- Transaction Hash: 4-Character Prefix + Suffix
- Counterparty: Full EOA Destination

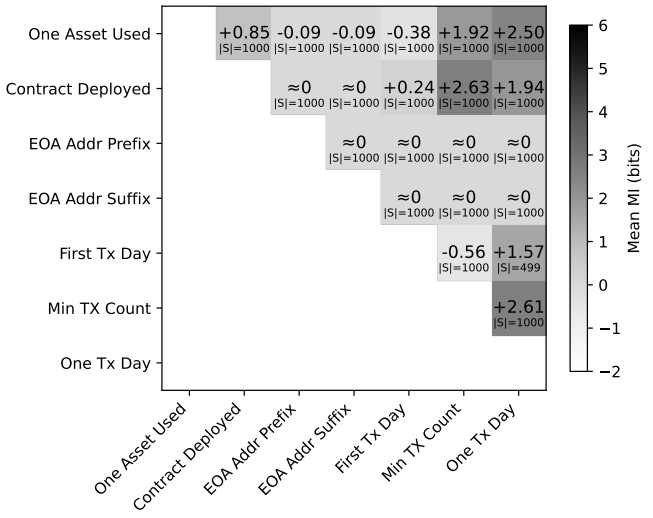
For *Counterparty: Full EOA Destination*, grouping by every unique Ethereum address scales to millions of groups. Prefix–suffix matching is similarly expensive: 4-character prefixes yield $16^4 = 65,536$ combinations, while combining prefix and suffix explodes to $16^8 \approx 4.3$ billion. To keep analyses tractable, we applied deterministic random sampling. Using BigQuery's `FARM_FINGERPRINT` function with a modulo operation, we selected 1% of addresses, ensuring reproducibility while reducing computational overhead.

Results. Our evaluation of individual cues confirms that the uniqueness and exactness of a cue instance can reduce anonymity significantly. The de-anonymization strength of a cue is primarily guided by these factors, making highly specific and rare cue instances far more effective than general or common ones.

The evaluation shows that cues derived from hashes provide exceptionally high de-anonymization potential. For instance, knowing eight characters from an EOA address (the first four and last four) provides an average Information Gain (IG) of ≈ 27.69 bits and



(a) Mean joint IG ($\overline{IG}_{\text{joint}}$) with Δ vs diagonal entries that show a single-cue IG. Darker shades indicate higher joint IG.



(b) Mean Mutual Information ($IG_{\text{indep}} - IG_{\text{joint}}$) and sample size $|\mathcal{S}|$. MI values: positive means overlap, negative means synergy.

Figure 5: Mean joint information gain ($\overline{IG}_{\text{joint}}$, bits) for cue pairs. Among single cues, *First Transaction Day* has the highest baseline IG in our snapshot, and its pairings achieve the largest mean joint IG (up to ≈ 19.20 bits, i.e., ≈ 359 candidate EOAs on average). Mean mutual information peaks at 2.63 bits for *Min TXs* $\times$ *Deployed*, indicating some redundancy. By contrast, two-character EOA address prefix/suffix cues are near-independent ($MI \approx 0$), and *Min TXs* $\times$ *First Tx Day* being synergistic ($MI < 0$) thereby improving pairwise IG. The sample size for *First Tx Day* $\times$ *Tx Time* is $n = 499$ because we average only over samples with nonempty joint support ($k_\cap > 0$) under randomized cue assignment.

is able to uniquely identify most accounts, with an Identification Rate (IR) of 99.8%. Similarly, knowing the first and last four characters of a transaction hash is sufficient to identify the associated EOA in 99.69% of cases.

For temporal cues such as Transaction Time, smaller and more exact time frames provide more information, as expected. Knowing the Hour of a transaction (avg. IG $\approx$ 13.91 bits) is more effective than knowing the Day (avg. IG $\approx$ 9.82 bits), which is in turn more effective than knowing the Month (avg. IG $\approx$ 5.79 bits).

The effectiveness of asset cues depends largely on uniqueness. For the Asset Transferred cue, transactions with rare assets (5k–10k users) yield the highest information gain (avg. IG $\approx$ 14.93 bits), followed by Uncommon assets (10k–100k users, avg. IG $\approx$ 13.07 bits) and Common assets (>100k users, avg. IG $\approx$ 8.36 bits). Numeric cues underscore the role of exactness. Knowing the Exact Quantity of a transfer is highly effective, achieving an Identification Rate of 70% (avg. IG $\approx$ 24.66 bits). Yet effectiveness varies widely, as non-round quantities are often unique, while common round numbers add little identifying power.

Overall, these individual cue evaluations show that instances with high uniqueness and/or exactness (exact values, specific addresses, rare assets, precise timing) yield significantly higher information gain than general characteristics (common asset usage, broad time ranges, low activity thresholds). Hash and numeric-based cue types (e.g., EOA Address, Asset Balance) had high identification rates in scenarios with high exactness. In contrast, categorical cue types (e.g., Asset Used) were inherently limited.

5.2 Combined Cues

The individual cue evaluation demonstrated that some cues individually can be very effective on their own, while others lack effectiveness. This section presents our evaluation of combining pairs of weaker cues. We generate 1,000 random data points for seven different cues (One Asset Used, Contract Deployed, 2 Character EOA Address Prefix and Suffix, First Tx Day, Min TX Count (fixed to 10) and One Tx Day), and evaluate their combinations. For the asset used, we select six widely used ERC-20 tokens: USDT, USDC, UNI, LIDO STETH, WBTC, and SHIB. These tokens have a larger user base and thus should yield lower individual identifiabilty. We measure both their Joint Information Gain, $IG_{joint}(C_0, C_1)$, and their Mutual Information ($MI(C_0, C_1)$), as defined in Equation 9. The sample size for *First Tx Day* $\times$ *Tx Time* is $n = 499$ because we average only over samples with nonempty joint support ($k_\cap > 0$) under randomized cue assignment. The results are nonetheless informative.

Results. The results are visualized in Figure 5. They show that combining cues, even those individually weak, can significantly enhance de-anonymization effectiveness. This is especially true when the combined cues are independent, meaning they provide distinct information and therefore have low mutual information. We can observe that combining cues leads to substantially higher IG than individual cues alone. For example, combining knowledge of the `First Transaction Day` (avg. $IG \approx 11.20$ bits) with the knowledge of a single commonly `Asset Used` (avg. $IG \approx 6.32$ bits)

yields a combined *IG* that is a little larger (0.38 bits) than either cue provides in isolation.

The effectiveness is related to the cues mutual information. We found that cues providing independent information, such as a partial `EOA Address`, consistently show very low to zero *MI* when paired with other cue types. This indicates high independence, meaning their de-anonymization power is almost perfectly additive. While some cues exhibit strong dependencies. For instance, knowing an EOA has used USDT makes it more likely it has also used USDC, resulting in high *MI* and a less than additive de-anonymization effect.

These findings demonstrate that an adversary does not need a single, highly effective cue. But a set of several, weaker cues can just as well or even more effectively de-anonymize a blockchain user. This is especially true, if the cues posed by the adversary have low mutual information.

6 Case Studies

To illustrate the real-world implications of cue-based de-anonymization, we found that specific case studies can lead to very high de-anonymization effects and identification rates. In this section, we begin with a simulation involving the Celsius data leak, refraining from actually using the real dataset to maintain ethical standards. We then focus on combining knowledge of multiple activity dates as well as knowledge of multiple token assets being used.

6.1 Celsius Data Leak

The Celsius data leak is a real-world example of off-chain cues being linked to real identities. During the company's bankruptcy case, documents were released that exposed customer names alongside transaction details such as dates, assets, and exact quantities. This presents a scenario where an adversary possesses a dataset of high-credibility cues linked to verified identities. We simulated an attacker attempting to leverage this information to identify the on-chain EOAs of Celsius depositors, by analyzing historical ERC20 token transfers sent to known Celsius hot wallets. Specifically, we simulate the leak by treating the transaction date, asset, and exact quantity as off-chain cues associated with each depositor, without using the leaked customer names themselves, and apply our matching procedure to determine how many EOAs are consistent with these cues.

We identify accounts who deposited funds to 11 confirmed Celsius Network wallet addresses (extracted from Etherscan's labeled accounts). The evaluation focuses on incoming transfers (from source address to Celsius wallets). We group this subset of users separately and incrementally by date, asset and quantity transferred, and determine how many unique EOAs are identifiable for each combination.

Our simulation revealed high risk, demonstrating that precise transaction details are sufficient to almost completely de-anonymize user pseudonymity. As shown in Table 4, an attacker with knowledge of just three cues from the leak: the transaction day, asset, and exact quantity, could potentially achieve a unique identification of over 85% of depositors (IR $\approx$ 85.7%).

Table 4: Celsius Leak simulation results: Celsius depositor account + up to 3 cues. Even a small set of leaked transaction details (day and exact quantity) is enough to identify the vast majority of depositors, highlighting the severe privacy risks of such leaks.

Cues Used	Avg. IG	Avg. k	IR (%)
(Depositor) + Day	18.19	723.26	0.01
(Depositor) + Asset	17.29	1348.60	0.09
(Depositor) + Quantity	27.21	1.39	74.98
(Depositor) + Day + Asset	21.97	52.38	0.60
(Depositor) + Day + Quantity	27.49	1.14	84.83
(Depositor) + Asset + Quantity	27.27	1.33	76.82
(Depositor) + Day + Asset + Quantity	27.51	1.13	85.68

The results clearly show that the `Exact Quantity` transferred is the dominant factor in this de-anonymization scenario. Knowing the quantity alone can uniquely identify nearly 75% of all depositors, with an average remaining anonymity set (k) of just 1.39. In contrast, knowing only the `Day` and/or `Asset` provided only a very small identification rate of less than 1%.

This case study clearly demonstrates the severe privacy risks of data leaks from centralized services. Our simulation shows that an attacker, using transaction details (like those exposed in the Celsius leak) alongside the customer names (also present in the leak), can theoretically (and thus practically) link real-world identities to the vast majority of user EOAs.

6.2 Knowledge of Activity Dates

This case study investigates the de-anonymization potential from knowing a set of specific dates on which a target EOA was active (sent at least once transaction that date). An attacker might gain this information by correlating various off-chain cues to a timeline or from data leaks that contain multiple transaction dates.

To focus the analysis on typical, non-automated users, we filtered out high-frequency accounts such as bots or exchange infrastructure (EOAs with > 250 transactions in 2024). The experiment assumes an attacker knows d distinct dates (where d ranges from 1 to 10) on which the target EOA initiated at least one transaction during the year 2024.

The results in Table 5 show a reduction in the anonymity set as the number of known activity dates increases. While knowing just 5 activity dates reduces the average anonymity set (k) to approximately 84 candidates, knowing 7 dates makes unique identification possible (IR $\approx$ 1.7%). With knowledge of 10 distinct activity dates, the average anonymity set is reduced to just 2.07 addresses, and the attacker can uniquely identify the target EOA in over 59% of the tested combinations. This case study demonstrates that a user's activity timeline creates a unique "fingerprint." The effectiveness increases with each additional date known.

6.3 Knowledge of Token Used

This case study investigates the de-anonymization potential of an adversary with knowledge of a set of specific tokens a user has used (sent at least once). An attacker might obtain such a set from

Table 5: De-anonymization results for known activity dates (Year 2024, EOAs with < 250 TXs). Even a handful of known transaction dates can drastically shrink anonymity sets, and with 10 dates an attacker can uniquely identify targets in the majority of cases.

Known Dates (d)	Avg. IG	Avg. k	IR (%)
1	9.23	361,082.20	0.0
2	13.91	14,082.09	0.0
3	16.84	1,839.90	0.0
4	19.16	369.76	0.0
5	21.29	84.14	0.0
6	22.87	28.14	0.0
7	24.31	10.44	1.7
8	25.55	4.39	22.6
9	26.28	2.66	42.6
10	26.64	2.07	59.2

Table 6: De-anonymization results for known token usage. Even knowledge of a handful of common tokens in a user's portfolio can sharply reduce anonymity, with 5 tokens often narrowing candidates to a few dozen and 10 tokens enabling unique identification in over half of cases.

Known Tokens (t)	Avg. IG	Avg. k	IR (%)
1	9.65	268,406.91	0.0
2	16.68	2,067.78	5.2
3	20.21	178.40	19.1
4	21.92	54.60	27.5
5	22.80	29.61	25.5
6	23.59	17.12	33.8
7	24.17	11.49	39.3
8	24.04	12.52	38.6
9	24.34	10.19	39.8
10	25.84	3.61	52.2

various sources, including a user's publicly shared portfolio, data leaks from exchanges or portfolio trackers, or by correlating the user with specific crypto-asset communities.

To focus the analysis on common cryptoassets and avoid identification through rare assets, we filtered the potential token list. Only tokens that had been used (sent at least once) by more than 50,000 unique EOAs before the cutoff (2025) date were considered candidates for the attacker's knowledge set (to avoid unused or rarely used assets). The experiment assumes an attacker knows t distinct tokens (where t ranges from 1 to 10) that a target EOA has sent at least once.

The results show, that by knowing just 5 common tokens used by an EOA, the average remaining anonymity set (k) reduces to approximately 30 candidates, with over 25% of cases leading to a unique identification. With knowledge of 10 common tokens, the average anonymity set is reduced to fewer than 4 addresses (avg. $k \approx 3.61$), and the attacker can uniquely identify the target EOA in over 52% of the tested combinations.

This case study demonstrates that knowing a user's portfolio, even when composed of relatively common tokens, is a highly effective de-anonymization attack. The cumulative effect of combining these cues drastically reduces the search space.

7 Prevalence of Cues on Social Media

Off-chain cues can be present on social media. Figure 6 shows four tweets in which users reveal partial information of their on-chain activity, which could be used to identify their accounts on a blockchain. The issue extends to a variety of platforms, including video content on Youtube, for example. Some users intentionally mask certain identifiers, but leave partial prefixes or suffixes visible, sometimes shown in wallets or block explorers, or simply disclose which assets they hold. In this section, we focus on the automated extraction and analysis of textual cues from posts on X (formerly Twitter) using Large Language Models (LLMs).

7.1 LLM-Based Cue Type Extraction

To enable an analysis of cue prevalence on social media posts, we further developed an LLM-based cue extraction methodology. This system uses a commercial state of the art LLM, specifically `gemini-2.5-pro-preview-05-06`, guided by a system prompt, which included the full cue taxonomy, task instructions, and examples to ensure consistent and accurate output formatting. The LLM is tasked to perform the following steps:

(1) Identify phrases in a text corresponding to the predefined cue types from our taxonomy.
(2) Extract the specific instance associated with each identified cue.
(3) Group related cues that likely refer to the same on-chain event or account.
(4) Assess credibility by assigning a rating based on the context and language of the post.
(5) Output the extracted information in a structured, parseable format for analysis.

We employ multishot prompting by providing examples of how a human would extract cues from tweets as part of the system prompt, as this has improved results significantly.

This automated methodology provides a scalable method of processing large volumes of textual off-chain data, which was used for our analysis of cue prevalence on social media sites.

7.2 Evaluation & Dataset Setup

We measure cue prevalence on social media using the publicly available Twitter dataset by Sawhney et al., from their paper "Cryptocurrency Bubble Detection: A New Stock Market Dataset, Financial Task & Hyperbolic Models" [18]. The dataset contains cryptocurrency related posts on X (formerly Twitter) between January 1, 2020 and April 7, 2021 and includes only tweets that mention at least one cashtag (e.g., $BTC, $ETH). From this dataset we sample 10,000 tweets uniformly at random, without additional filters for analysis, and analyzed them in chunks of 10 posts. We did not apply extra filters for bots, spam or scams.

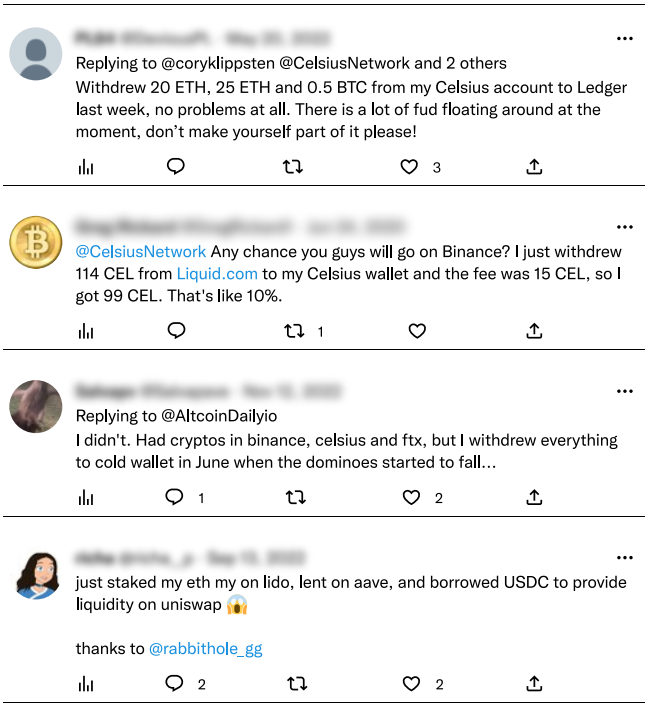


Figure 6: Four tweets in which users reveal partial information of their on-chain activity, which could be used to identify their accounts on the blockchain.

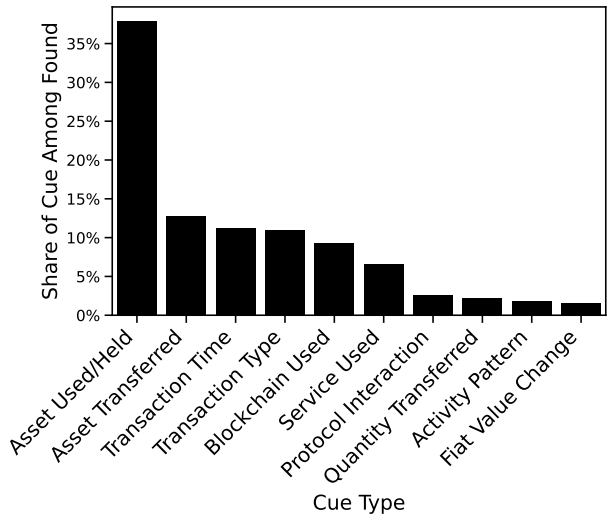


Figure 7: Distribution of cue types among all 8,647 found cues within 2,381 tweets. Only showing cue types with more than 1%. Multiple cues can be present within a single Tweet. Asset-related cues were the most commonly found, making up almost 50% (first two bars).

7.3 Results for Cue Prevalence

Our analysis revealed a high prevalence of cues in cryptocurrency-related postings on X. Out of the 10,000 tweets analyzed, 23.81% were found to contain at least one identifiable off-chain cue. This suggests that approximately one in four cryptocurrency-related tweets may contain potentially de-anonymizing information. Among these 2,381 tweets, we identify 8,647 cue types, which means a single tweet on average contains 3.63 cues.

As shown in Figure 7, cues related to assets were the most common, with `Asset Used/Held` and `Asset Transferred` together accounting for nearly 50% of all identified cues. This reflects a tendency of users to explicitly mention which tokens they hold or move, often in casual or promotional contexts. In contrast, cues we identified as having very high de-anonymization potential on their own, such as `EOA Address` or exact `Asset Balance`, were observed with much lower frequency. The relative rarity of such precise disclosures reduces the prevalence of direct leakage of highly sensitive identifiers, but does not eliminate the risks associated with more indirect information.

In line with our previous finding that even seemingly weaker cues carry significant analytical value when aggregated (c.f. Section 5.2, knowledge of asset usage patterns, transaction timing, or service interaction may not reveal a unique identity in isolation, but in combination they may significantly narrow the anonymity set of a user. Similarly, as demonstrated in our "Knowledge of Token Usage" case study (6.3), combining these common, individually weaker cues can still have a significant cumulative risk to user pseudonymity. Overall, this analysis suggests that while direct identifiers appear rare, off-chain discourse provides a rich set of cues that adversaries could exploit, particularly around knowledge of a portfolio of tokens.

7.4 LLM Effectiveness Evaluation

We evaluate the accuracy of the cue extraction prompt using two ground-truth datasets: a synthetically generated dataset and a manually labeled dataset. The evaluation covers two tasks: (i) detecting whether a post contains any off-chain cue from our taxonomy, and (ii) identifying the cue types present when cues exist.

Evaluation Tasks and Metrics. Let $y \in \{0, 1\}$ denote cue presence in a post and $\hat{y}$ the model prediction.

Cue Presence Detection. We treat cue presence as a binary classification task and report accuracy.

Cue-Type Inclusion (Type Recall). For posts containing cues, we measure how often the extractor includes each ground-truth cue type in its output and report micro-averaged recall across all cue-bearing posts.

7.4.1 Synthetic Dataset Evaluation. We first evaluate the cue extractor on a controlled synthetic dataset of 300 posts generated using LLMs. To enable a balanced assessment, 150 posts contain at least one cue from our taxonomy, while the remaining 150 are cryptocurrency-related but cue-free. All posts were manually reviewed to establish ground-truth labels. To reduce labeling ambiguity, the cue types `Protocol Interaction` and `Service Used` are merged for the purpose of extractor evaluation. We compare three Google Gemini models (2.0-flash, 2.5-flash, and 2.5-pro). The generation prompt used to create the synthetic dataset is included in the source code repository.

Table 7: Cue extractor performance on the synthetic dataset. All models reliably detect cue presence; type-level inclusion recall improves substantially with larger models.

Model	Avg. Time	Cue-Presence Acc.	Type-Recall (Incl.)
2.0-flash	0.6 sec	98.33% (295/300)	55.33% (83/150)
2.5-flash	2.1 sec	98.33% (295/300)	80.00% (120/150)
2.5-pro	11.2 sec	99.00% (297/300)	86.00% (129/150)

LLM Performance Evaluation Results. The performance of the evaluated models (Gemini-2.0-flash, Gemini-2.5-flash-preview-0520, and Gemini-2.5-pro-preview-0506) is summarized in Table 7. During evaluation, posts were processed in batches of ten tweets per model invocation to reflect the configuration used in the main extraction pipeline.

7.4.2 Manually Labeled Dataset Evaluation. To assess performance under more realistic and potentially ambiguous conditions, we additionally evaluated the extractor on a manually labeled dataset. We randomly sampled 300 posts from the cryptocurrency tweet dataset by Sawhney et al. and annotated them according to our cue taxonomy. This dataset was evaluated using Gemini-2.5-pro.

On this dataset, the model achieved an accuracy of 88.33% (265/300) in distinguishing posts with and without cues. For posts containing cues, the minimum cue-set coverage accuracy was 30.43% (7/23), indicating that only a third of the human-annotated cue sets were fully included in the model output.

These results suggest that cue *presence* detection remains largely consistent with human judgment, while cue *type* extraction is very sensitive to annotation differences. A qualitative comparison reveals that discrepancies primarily arise from differences in implicit assumptions rather than outright extraction failures. In particular, human annotators sometimes:

- inferred the underlying blockchain from the mentioned asset (e.g., assuming Ethereum for ERC20 tokens), whereas the LLM required explicit evidence,
- assumed asset ownership based on positive or promotional statements about an asset, which the LLM did not interpret as a cue,
- overlooked cues that were detected by the LLM and found to be valid upon closer manual inspection.

Overall, the synthetic evaluation shows that all tested models reliably distinguish posts with cues from those without, while cue-type inclusion improves with model capacity at the cost of increased inference time. The manual evaluation confirms that cue presence detection remains largely consistent with human annotation, but also reveals substantially lower agreement at the cue-type level under real-world conditions. These discrepancies are primarily attributable to differing implicit assumptions made by human annotators versus the extractor's reliance on explicit evidence. Taken together, the results support our choice of the Gemini-2.5-pro model for empirical cue extraction, while indicating that type-level coverage should be interpreted conservatively when applied to social media data.

8 Discussion

The evaluation presented in the previous section quantified the effectiveness of using various off-chain cues in different scenarios to reduce the anonymity set of blockchain users, specifically on Ethereum. The results from evaluating individual cues, their combinations, and realistic case studies provide strong evidence that even partial off-chain information can significantly compromise user pseudonymity on public blockchains. Furthermore, the investigation into social media showed a concerning prevalence of such cues. This section discusses these findings, their implications, and acknowledges the limitations of this study.

8.1 Implications

The underlying Data Type of Cues Sets the De-anonymization Potential. The underlying cue data type determines the inherent de-anonymization potential, even before the attributes of a specific cue instance may be taken into account. Hash-based cues (e.g., EOA addresses) inherently have the highest potential, as they are designed for uniqueness within a vast value space. Numeric cues (e.g., quantity transferred, asset balance) also have a large value space, giving them high potential for de-anonymization. However, their practical effectiveness is often limited because instances of numeric cues frequently cluster around common values (e.g., 1 ETH, 100 USDT). Similarly, temporal cues (e.g., transaction time) can be effective, but their effectiveness is limited when they correspond to periods of high on-chain activity. Finally, categorical cues (e.g., asset used, protocol interaction) are inherently limited by the popularity of specific choices within that set.

Uniqueness and Exactness Drive De-anonymization. A primary insight from the individual cue analysis is the importance of uniqueness and exactness. Cue instances corresponding to statistically rare on-chain events (e.g., non-rounded amounts with multiple decimal precision, interactions with rare tokens, exact hashes) show strong de-anonymization effectiveness. In many instances, such cues were capable of reducing anonymity sets by several orders of magnitude, enabling unique identification. Conversely, cues related to common activities, such as transferring common assets (e.g., USDT) or interacting with popular DeFi protocols during broad timeframes, were individually much weaker, as the target EOA is effectively hidden within the set of other EOAs performing the same action.

Combining Cues Increases Effectiveness. While individual cues varied in their de-anonymization strength, the combination of multiple cues consistently increased their combined effectiveness. Our findings showed that even two or three cues, especially those with low mutual information, could drastically narrow the search space far more effectively. For example, knowing a blockchain user was active on a few specific days, without knowing other transaction details, could create a unique "fingerprint" that might be capable of identifying the user's EOA. This shows that effective de-anonymization does not necessarily require highly exact or rare individual cues, but a combination of several less effective cues can be equally or even more effective.

Case Studies Confirm High Risk. The case studies, in particular the one based on the Celsius data leak, demonstrate that with a few exact data points (date, asset, quantity) a vast majority of users could

be de-anonymized and identified with high confidence. Similarly, knowing a small set of common tokens used by an EOA, or a set of its activity dates, drastically reduced anonymity, often to a level for targeted analysis. These findings indicate that compromising blockchain pseudonymity using off-chain cues is often feasible and requires less information than perhaps commonly perceived.

"Noise" from High-Activity and Automated Accounts. The evaluation revealed that high-frequency EOAs, bots, exchange wallets, or oracle services with high transaction counts, can act as "noise" in the de-anonymization effects, especially for cues related to general activity. Filtering these entities, as done in, e.g., the case study on knowledge of activity dates (c.f. Section 6.2, is crucial for assessing the de-anonymization risk for typical users. Without filtering, some scenarios involving multiple cues of the same type would converge against the set of these High-Activity EOAs and the Mutual Information will increase with each cue added.

Effectiveness of the LLM-Based Cue Extractor. Automated cue extraction using LLMs lowers the barrier for systematically collecting de-anonymizing information from public text, increasing the feasibility of large-scale linkage attacks. Manual evaluation indicates that cue presence detection aligns well with human judgment, while cue-type extraction is more sensitive to implicit assumptions.

Notable Cue Prevalence on Social Media. The findings from this social media analysis of tweets on X, demonstrated that cues can be found much more frequently than commonly thought, with 23.81% of analyzed tweets (2,381 out of 10,000) containing at least one identifiable off-chain cue. The prevalence of asset-related cues (Asset Used/Held, Asset Transferred) together making almost 50% of all cues identified, indicates that users frequently discuss their holdings and transactional activities. While these cues are often weak individually, we find that combining multiple weaker cues can, in some scenarios, effectively de-anonymize a blockchain user.

Implications for User Privacy and Security. Overall, the findings demonstrate that blockchain pseudonymity is easily compromised by off-chain cues. Users must be cautious when sharing details about their transactions, holdings, or activity patterns. Data leaks containing links between real-world identities to on-chain activity, such as the Celsius leak, are especially dangerous to user privacy when combined with public blockchain analysis. This work highlights the need for better user education regarding privacy best practices and potentially provides a motivation for the development and adoption of more robust on-chain privacy solutions or privacy-preserving off-chain communication methods.

8.2 Limitations

The findings presented in this paper, while demonstrating the potential for de-anonymizing blockchain users using off-chain cues, must be considered within the context of specific limitations.

Ethereum-Specific Focus. Our evaluation was exclusively using data from the Ethereum blockchain. Ethereum's characteristics of its on-chain data (e.g., transaction patterns, smart contract usage, EOA behavior) may differ significantly from other blockchains, especially non-EVM chains like Bitcoin or Solana, or privacy-focused chains like Monero. Therefore, the results regarding the effectiveness of certain cues may not generalize to other blockchains. However, the underlying principles, that unique and exact cues are effective, and combining cues enhances de-anonymization, are likely to generalize across all non-privacy-focused blockchains.

Attributes of Cue Instances in Evaluation. While we tested variations in uniqueness and exactness for some cues (e.g., exact amounts versus ranges, common vs rare tokens), the cues used in our analysis were assumed to be true representations of actual on-chain activities or account characteristics. In real-world scenarios, especially with cues from social media, cues can be exaggerated, deliberately misleading, or entirely fabricated. Our current methodology does not model the impact of such low-credibility cue instances, which could reduce the chance of a successful de-anonymization.

Limitations of Cue Extractor. The LLM-based cue extractor is effective but has inherent limitations. Due to the probabilistic nature of LLMs, its outputs may contain errors, and extraction is restricted to cue types defined in the taxonomy, excluding previously unseen cues. In addition, the current implementation operates only on textual data and cannot process visual content, such as screenshots of wallet balances.

Limitations of Social Media Analysis. The dataset was sourced from X (formerly Twitter) with posts from January 2020 to April 2021. Posting behaviors and cue prevalence may differ on other platforms (e.g., Reddit, Discord, Telegram) or time periods. Secondly, the dataset only contains posts with at least one "cashtag." This excludes posts without explicit cashtags, which might have a different distribution and nature of cues. Further, the analysis did not include a filtering for scam, spam, or bot-generated posts. Therefore, the reported prevalence of cues might be inflated by non-genuine posts, as these could contain fabricated or misleading information designed to mimic real user activity.

8.3 Potential Countermeasures

The de-anonymization risks quantified in this paper is significant. Mitigating these risks can be done from different levels. We outline several potential countermeasures below.

- **User-Level** Since some cues are voluntarily disclosed by blockchain users, the primary mitigation can be user education on privacy best practices. This could be avoiding the sharing of exact transaction amounts, and could be improved by tools, such as a "Content Risk Analyzer", that scans social media posts for high-risk cues before or after they are published.
- **Application-Level** Applications such as wallets and exchanges can be designed to promote better privacy practices. This could include warnings before users share sensitive data like transaction hashes (e.g, when a user copies a hash) and providing tools for managing multiple EOA.
- **Blockchain-Level** The most robust mitigation is the use of on-chain privacy technologies. Protocols such as mixers (e.g., Tornado Cash) make linkage attacks impossible and specific privacy blockchains (e.g., Monero) make these kind of linkage attacks completely impossible due to the built-in privacy features.

8.4 Ethical Considerations

We study de-anonymization to expose risks and strengthen defenses, not to enable attacks. We've analyzed only public Ethereum data (via Google BigQuery) and a public Twitter research dataset. No scraping or access circumvention were used by us. For the Celsius case, we did not process the actual disclosed spreadsheets. We only simulated their effect and derivatives of it. We did not contact or target individuals and performed no active probing. All analyses have been performed offline. Our LLM use is limited to extracting cue types from text, not to directly de-anonymize users.

9 Related Work

In this section, we examine the literature related to the core challenge of this paper. We first review de-anonymization techniques, then explore their application to blockchains, and finally discuss prior research that connects social media with on-chain activity.

9.1 De-anonymization Attacks

The work of Narayanan and Shmatikov is widely regarded as the foundational work in de-anonymization via linkage attacks [14]. Their key finding was that even without explicit identifiers, individuals can be re-identified by matching a small number of unique data points against auxiliary public information. This principle, that sparse datasets are especially vulnerable because users exhibit unique patterns of behavior, forms the theoretical basis of our investigation.

9.2 De-anonymization Attacks on Blockchains

One of the earliest papers exploring anonymity for blockchains was by Reid and Harrigan [17]. Their work analyzed on-chain transaction graphs and demonstrated how "off-network information," such as Bitcoin addresses posted on forums, could be used to link real-world identities with on-chain activity. Much of the related work labeled "de-anonymization" in blockchain research refers to techniques for address clustering, with multiple works on Bitcoin [10, 12, 13] and Ethereum [3, 21]. These methods primarily rely on analyzing on-chain transaction patterns to group addresses belonging to the same entity, rather than linking them to a specific off-chain identity.

Especially relevant to this paper is the work by Zhang et al., who investigated the de-anonymization of Litecoin users making online payments [23]. Their methodology used web trackers (cookies) to capture off-chain purchasing information (price, time) and performed transaction linkage attacks to correlate this data with on-chain transactions, having high success rates in their simulations. While these studies demonstrated the potential of using off-chain data for linkage attacks, our work provides a comprehensive taxonomy of the diverse cues available and quantifying their de-anonymization potential for the smart contract platform Ethereum.

9.3 Linking Social Media to On-Chain Activity

Research directly linking specific off-chain social media data to on-chain activity is closely related to our work. Most of this research has focused on social media platforms, especially on X ("Crypto Twitter") and primarily on analyzing macro activities. Studies by Kang et al. and Sawhney et al. have successfully correlated millions of crypto-related posts on X (formerly Twitter) with on-chain events, demonstrating that spikes in public sentiment often precede major market movements or real-world events [9, 18]. These studies demonstrate that public off-chain data can reflect and even predict on-chain activities. Our work builds on this finding, but focuses on individual users, instead of global sentiment analysis, demonstrating how cues from individual posts can be used to directly de-anonymize their authors.

10 Conclusion and Future Work

This paper investigated the potential for identifying blockchain users by correlating publicly available off-chain cues with on-chain data. To this end, we developed a structured framework, creating a comprehensive taxonomy to classify off-chain cues, proposing matching algorithms to link them to on-chain activity, and developing an LLM-based extractor to automate cue discovery from unstructured text. Our large-scale empirical evaluation on the Ethereum blockchain systematically quantified the de-anonymization risk posed by these cues.

Our key findings demonstrate that the pseudonymity of public blockchains is vulnerable when cues are leaked. The effectiveness of a cue is directly related to its uniqueness and exactness, with specific, non-round transaction quantities and interactions with rare assets being exceptionally powerful. We showed that even a small combination of 2 to 3 moderately specific cues can dramatically reduce the anonymity set, in some cases to unique identification. This was illustrated in real-world case studies: a simulation of the Celsius data disclosure revealed that over 85% of depositors could be identified, and our social media analysis found that approximately 24% of cryptocurrency-related tweets contained potentially de-anonymizing cues. While these are specific instances, and cue-type interpretation can be ambiguous under real-world conditions, cue presence is detected robustly, and the overall de-anonymization risk depends primarily on an attacker's ability to obtain off-chain cues.

In conclusion, this work has systematically demonstrated the practical feasibility and risk of linking off-chain cues to on-chain data for user identification. Our results underscore that blockchain pseudonymity is *at risk* when off-chain cues are revealed, highlighting the need for better privacy practices. Potential future directions include integrating these off-chain methods with on-chain heuristics like address clustering, extending the analysis to other blockchains, and developing practical tools to help users assess and mitigate their de-anonymization risk before sharing information publicly. For example, for user-side mitigation, one could build a browser extension or wallet plugin (a *Content Risk Analyzer*) that scans outgoing messages and flags potential cues (e.g., warning if an exact crypto amount is posted).

Acknowledgments

Research supported by the German Research Foundation (DFG), Schwerpunktprogramm SPP 2378 - ReNO-2 (511099228), 2025-2029. The authors used generative AI-based tools including ChatGPT 5 to revise the text, improve flow and correct typos, grammatical errors, and awkward phrasing.

References

[1] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. 2013. Evaluating user privacy in bitcoin. In *International conference on financial cryptography and data security*. Springer, 34–51.

[2] Raphael Auer, Bernhard Haslhofer, Stefan Kitzler, Pietro Saggese, and Friedhelm Victor. 2024. The technology of decentralized finance (DeFi). *Digital Finance* 6, 1 (2024), 55–95.

[3] Ferenc Béres, István A Seres, András A Benczúr, and Mikerah Quintyne-Collins. 2021. Blockchain is watching you: Profiling and deanonymizing ethereum users. In *2021 IEEE international conference on decentralized applications and infrastructures (DAPPS)*. IEEE, 69–78.

[4] Wren Chan and Aspen Olmsted. 2017. Ethereum transaction graph analysis. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*. 498–500. https://doi.org/10.23919/ICITST.2017.8356459

[5] Weili Chen, Xiongfeng Guo, Zhiguang Chen, Zibin Zheng, and Yutong Lu. 2020. Phishing Scam Detection on Ethereum: Towards Financial Security for Blockchain Ecosystem.. In *IJCAI*, Vol. 7. 4456–4462.

[6] Mauro Conti, E. Sandeep Kumar, Chhagan Lal, and Sushmita Ruj. 2018. A Survey on Security and Privacy Issues of Bitcoin. *IEEE Communications Surveys & Tutorials* 20, 4 (2018), 3416–3452. https://doi.org/10.1109/COMST.2018.2842460

[7] Steven Goldfeder, Harry Kalodner, Dillon Reisman, and Arvind Narayanan. 2017. When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. *Proceedings on Privacy Enhancing Technologies* 2018 (08 2017). https://doi.org/10.1515/popets-2018-0038

[8] Harvard Law Review. 2025. Data Privacy in Bankruptcy: The Consumer Privacy Ombudsman. *Harvard Law Review* (March 2025). https://harvardlawreview.org/print/vol-138/data-privacy-in-bankruptcy-the-consumer-privacy-ombudsman/ Accessed: 2025-05-04.

[9] Inwon Kang, Maruf Ahmed Mridul, Abraham Sanders, Yao Ma, Thilanka Munasinghe, Aparna Gupta, and Oshani Seneviratne. 2024. Deciphering Crypto Twitter. In *ACM Web Science Conference (Websci '24)*. ACM, 331–342. https://doi.org/10.1145/3614419.3644026

[10] George Kappos, Haaroon Yousaf, Rainer Stütz, Sofia Rollet, Bernhard Haslhofer, and Sarah Meiklejohn. 2022. How to peel a million: Validating and expanding bitcoin clusters. In *31st usenix security symposium (usenix security 22)*. 2207–2223.

[11] Jieli Liu, Jinze Chen, Jiajing Wu, Zhiying Wu, Junyuan Fang, and Zibin Zheng. 2024. Fishing for fraudsters: Uncovering ethereum phishing gangs with blockchain data. *IEEE Transactions on Information Forensics and Security* 19 (2024), 3038–3050.

[12] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. 2013. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*. 127–140.

[13] Malte Möser and Arvind Narayanan. 2022. Resurrecting address clustering in bitcoin. In *International Conference on Financial Cryptography and Data Security*. Springer, 386–403.

[14] Arvind Narayanan and Vitaly Shmatikov. 2008. Robust De-anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. 111–125. https://doi.org/10.1109/SP.2008.33

[15] Marilyne Ordekian, Gilberto Atondo-Siu, Alice Hutchings, and Marie Vasek. 2024. Investigating Wrench Attacks: Physical Attacks Targeting Cryptocurrency Users. In *6th Conference on Advances in Financial Technologies (AFT 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 24–1.

[16] Masarah Paquet-Clouston, Matteo Romiti, Bernhard Haslhofer, and Thomas Charvat. 2019. Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. In *Proceedings of the 1st ACM conference on advances in financial technologies*. 76–88.

[17] Fergal Reid and Martin Harrigan. 2011. An Analysis of Anonymity in the Bitcoin System. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. 1318–1326. https://doi.org/10.1109/PASSAT/SocialCom.2011.79

[18] Ramit Sawhney, Shivam Agarwal, Vivek Mittal, Paolo Rosso, Vikram Nanda, and Sudheer Chava. 2022. Cryptocurrency Bubble Detection: A New Stock Market Dataset, Financial Task & Hyperbolic Models. arXiv:2206.06320 [cs.CL] https://arxiv.org/abs/2206.06320

[19] Ruisheng Shi, Yulian Ge, Lina Lan, Zhiyuan Peng, Shenwen Lin, and Lin Li. 2024. Deanonymizing Transactions Originating from Monero Tor Hidden Service Nodes. In *Companion Proceedings of the ACM Web Conference 2024* (Singapore, Singapore) *(WWW '24)*. Association for Computing Machinery, New York, NY, USA, 678–681. https://doi.org/10.1145/3589335.3651487

[20] Taro Tsuchiya, Jin-Dong Dong, Kyle Soska, and Nicolas Christin. 2025. Blockchain Address Poisoning. *arXiv preprint arXiv:2501.16681* (2025).

[21] Friedhelm Victor. 2020. Address clustering heuristics for Ethereum. In *International conference on financial cryptography and data security*. Springer, 617–633.

[22] Kristin Weber, Andreas E Schütz, Tobias Fertig, and Nicholas H Müller. 2020. Exploiting the human factor: Social engineering attacks on cryptocurrency users. In *International Conference on Human-Computer Interaction*. Springer, 650–668.

[23] Zongyang Zhang, Jiayuan Yin, Yizhong Liu, and Jianwei Liu. 2020. Deanonymization of Litecoin Through Transaction-Linkage Attacks. In *2020 11th International Conference on Information and Communication Systems (ICICS)*. 059–065. https://doi.org/10.1109/ICICS49469.2020.239510

[24] Guy Zyskind, Oz Nathan, and Alex 'Sandy' Pentland. 2015. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *2015 IEEE Security and Privacy Workshops*. 180–184. https://doi.org/10.1109/SPW.2015.27