

From Patterns to Profiles

A Multi-Dimensional Profiling Analysis of User Behavior and Risk on a Seized Cryptocurrency Exchange

SEN2231: CoSEM Master Thesis

S.E. van der Boon

From Patterns to Profiles

A Multi-Dimensional Profiling Analysis of User Behavior and Risk on a Seized Cryptocurrency Exchange

by

Simone Elaine van der Boon

S.E. (Simone) van der Boon

Student ID: 5086620

In partial fulfillment of the requirements for the degree of Master of Science in *Complex Systems Engineering & Management* (CoSEM) at Delft University of Technology.

To be defended publicly on Wednesday June 17th, 2026 at 3:00 PM.

Committee:	Dr. Rolf van Wegberg (Chair & 1 st Supervisor)	TU Delft
	Prof. Dr. Martijn Warnier (2 nd Supervisor)	TU Delft
	Bas Maliepaard (External Supervisor)	FIOD
Project Duration:	December 2025 - June 2026	
Faculty:	Faculty of Technology, Policy & Management (TPM)	

Cover image by wirestock on Freepik

An electronic version of this thesis is available at <https://repository.tudelft.nl/>

Acknowledgements

Those who know me well know that I have long looked forward to writing this specific page. Perhaps more than any other part. I will leave it to you, as the reader of this thesis, to judge how serious that statement is. In my spare time, I like to read books and jot down a few lines of context on the first page of those I buy or receive: when I got them, why I chose them, or what I have learned from them. It serves as a small souvenir of time and context on the bookshelves of my future self. Logically, the same holds true for this thesis, as it marks the end of almost seven years I have spent as a student in Delft. For that reason, it holds value to me in more ways than one, and I would like to express my gratitude to wonderful people who contributed to this incredible chapter.

First and foremost, Rolf. Not only for supporting me during the writing of this thesis, but also for the past four years in which we have known each other. I vividly remember our first contact during my board year at S.V.T.B. Curius, and shortly thereafter the enthusiasm I felt when you gave me the opportunity to become a research assistant. Over the past four years, you have been a true mentor, teaching me not only on an academic level, but also about life — including all its challenges and joys. You always provided a listening ear, gave me the space to flourish, help discover my broad range of interests, find enjoyment and most important: self-belief. A radio debut, writing a book chapter, teaching courses within the bachelor curriculum, developing an educational kit on cybersafety for primary school children. These are just a few of so many highlights. It has always felt incredibly extraordinary to have a job like this next to my studies. You have, without a doubt, left a mark on my personal development in the best possible way during my years as a student. Thank you for everything.

Bas, thank you for being such a great external supervisor and giving me insights into your cool job. Your down-to-earth way of thinking, helpfulness and endless patience in explaining the world of cryptocurrencies were invaluable in completing this thesis with a big smile. Sharing your enforcement perspective (often combined with a serious lack of sleep), gave this thesis its real-world value.

Lastly, Martijn. Your enthusiasm during my kick-off has stayed with me ever since. I hope you realize how much your genuine interest inspires and motivates others. Thank you for your availability, your everlasting critical eye, and for giving me confidence where I sometimes struggled to find it myself.

Naturally, I also want to thank my colleagues at FACT for welcoming me in your team and involving me wherever possible in your exciting field of work. Not forgetting to mention, for introducing me to the (now not so secret) *snack-khuis*. All of you were always willing to discuss my thesis and share your enforcement perspectives, while also being open to good conversations, coffee across the street, push-up challenges and many laughs in between. Thanks to you all, the past few months have enriched both my academic and personal experience far more than I expected.

To my *cyber-cyber* colleagues: Cécile, Fieke, Kelvin, Maaïke, Pepijn, Hugo and naturally Joyce, thank you for always thinking along, offering your help, and sharing plenty of laughter along the way. I wish you every success in your academic journeys. A special note of appreciation also goes to Joy. Gerlofke, thank you for always checking in on me and guiding me when I needed it most. Your support is of immeasurable value to TPM students.

To my dear friends from the *Kleiweg*, *TVP*, the *Birthday Club*, 29, as well as Thomas, Evalie and my current roomie Janne: thank you for shaping my student years and for all the memories we got (and still get) to share. Lastly, mama, papa and my brother Bram: thank you for being my greatest source of support. Mama, for always being there to listen, cheering me on with every (stupid) thing I do, and for the literal survival packages. Bram, for always making me laugh and for putting me back with two feet on the ground, which is often more necessary than I care to admit. To my dear Opa 'Snor': I know you are extremely proud from up above. I'm finally an engineer, just like you!

I will end this shameless *lofzang* here. Have fun reading my thesis!

Simone van der Boon
Rotterdam, June 2026

Executive Summary

Cryptocurrencies play an increasingly prominent role in legal financial activities, while simultaneously creating new opportunities for illicit use. Their speed, ease of crossing borders, and pseudo-anonymous nature strongly contribute to their potential for laundering illegal funds. Cryptocurrency exchanges are considered critical financial services that convert cryptocurrencies to fiat money and vice versa, thereby fulfilling a central role within the crypto ecosystem. As a result, increased attention is directed towards such digital infrastructures that may facilitate and scale illicit financial activity. Gaining insights into the users operating on such platforms, as well as the behaviors they perform, may support law enforcement in understanding how these infrastructures are utilized and which behavioral patterns emerge. Moreover, the data obtained from these vital platforms enhances the potential of data-driven approaches that identify broader behavioral structures and aid differentiating between user groups, which can in turn support the prioritization of investigations.

Understanding user profiles and their behavioral patterns within the internal structures of cryptocurrency exchanges has received limited attention in earlier research, yet understanding (criminal) modus operandi is considered essential for effective intervention. Previous studies mostly rely on identifying suspicious activity through machine learning models, using publicly available blockchain data, often without including the context and perspectives of investigative practice. By examining a unique dataset from the cryptocurrency exchange *Cryptex* acquired through law enforcement operations in 2024, this study aims to fill this knowledge gap.

The central research question guiding this thesis is: *“To what extent can latent behavioral profiles be identified within the internal records of a seized cryptocurrency exchange, and what insights do these profiles provide for law enforcement investigations?”*

By applying a mixed-methods approach, this research conducted a Latent Profile Analysis (LPA) to identify distinct behavioral user profiles based on transactional activity indicators related to deposit and withdrawal operations on a seized cryptocurrency exchange.

First, a conceptual lens was developed based on behavioral indicators related to structuring transactions, informed by policy literature on suspicious transactional activity within cryptocurrency ecosystems. Several methodological selection criteria were established to assess the suitability of the empirical dataset for LPA specifically, after which a study population was defined, consisting only of a subset of users who performed both deposits and withdrawals within a limited timeframe.

Next, the obtained profiles were further enriched by analyzing their differences in variables associated with elevated money laundering risk, including withdrawal address-sharing patterns, illicit exposure of incoming transactions, together with adherence to Know-Your-Customer (KYC) regulations and the use of privacy coins, both of which are associated with increased attribution mitigation.

Finally, an exploratory focus group with law enforcement professionals was conducted to assess the relevance, interpretability, and potential investigative value of the obtained behavioral profiles and the methodological approach applied in this thesis.

In answering the main research question this study demonstrates that distinct latent profiles can be identified within cryptocurrency exchange data, reflecting distinguishable behavioral patterns and broader behavioral strategies within the exchange ecosystem. These profiles primarily differ in their activity intensity, variability in transaction sizes and the extent to which multiple cryptocurrencies were exchanged. High-activity profiles on *Cryptex* are associated with increased interaction on the blockchain, greater overlap with external wallet addresses, and higher exposure to illicit activity, suggesting more complex or potentially coordinated behavioral patterns. In contrast, lower-activity profiles exhibit limited on-chain interaction and appear more aligned with KYC-procedures, indicating towards potential alternative transaction strategies that are less visible within publicly available blockchain data.

These findings show that user behavior within cryptocurrency exchanges is not homogeneous, but can be meaningfully structured into distinct groups that capture relevant behavioral differences from a money laundering perspective. Rather than directly identifying illicit activity, the findings suggest

that behavior related to money laundering risk may emerge through broader behavioral configurations which are built upon combinations of transactional and platform-related indicators. While the profiles do not directly identify illicit activity, according to law enforcement professionals they appear to provide a useful abstraction for reducing complexity in large transactional datasets and can support exploratory analysis, prioritization, and hypothesis generation within law enforcement investigations. However, the method is considered unable to capture important nuance or answer cascading questions which are essential in law enforcement investigations, highlighting the remaining importance of human intelligence.

In doing so, this research provides three key contributions. First, it demonstrates the applicability of LPA as a method to uncover latent behavioral structures in cryptocurrency exchange transactional data, by means of a unique empirical dataset. This is considered to provide a more transparent approach to understanding user behavior and clustering. Second, this study demonstrates that cryptocurrency exchange users exhibit distinct behavioral configurations that become visible when transactional, blockchain-related, and risk-related indicators are analyzed in combination instead of isolation. The findings suggest that potentially elevated money laundering risk is therefore not uniformly distributed. Moreover, such risk does not emerge from a single behavioral dimension alone, but rather through broader combinations of interacting indicators across multiple behavioral dimensions. Therefore, the results suggest that blockchain-based indicators alone may not sufficiently capture the full diversity of transactional strategies, which potentially leaves certain *modus operandi* underrepresented or overlooked. This highlights the added value of combining internal exchange data with blockchain-related indicators for a more context-sensitive understanding.

From an investigative perspective, the findings demonstrate how behavioral profiling approaches may support exploratory analysis, data structuring, prioritization, and hypothesis generation within intelligence-driven cryptocurrency investigations. At the same time, the study emphasizes that the interpretation of such profiles remains highly dependent on contextual understanding, human expertise and nuance by investigators.

Contents

Acknowledgements	i
Executive Summary	ii
List of Figures	vi
List of Tables	vii
List of Abbreviations	ix
1 Introduction	1
1.1 Formal Takedown of Cryptex	2
1.2 Previous Research and Knowledge Gaps	2
1.2.1 User profiles and Behavioral Aspects Related to Cryptocurrency Exchanges	2
1.2.2 Access to Empirical Exchange-Related Data	3
1.2.3 Limited Inclusion of Law Enforcement Perspectives	3
1.3 Thesis Scope	3
1.3.1 Problem Statement	3
1.3.2 Research Objective	4
1.3.3 Research Questions	4
1.3.4 Link to the MSc CoSEM Programme	5
1.4 Thesis Outline	5
2 Technical and Contextual Background	6
2.1 Fundamentals of Bitcoin and Blockchain Technology	6
2.2 Illicit Financial Flows and Money Laundering	8
2.3 Cryptocurrency Exchanges as Criminal Facilitators	8
2.3.1 Illicit Potential across Cryptocurrency Exchanges Types	8
2.4 Common Money Laundering Practices	9
2.5 Investigative AML Tooling and Blockchain Intelligence	10
2.5.1 Analytical Approaches in Crypto AML	10
2.5.2 Chainalysis	11
3 Case description: Cryptex	12
3.1 Investigation by Law Enforcement	12
3.2 Services Offered	12
3.3 Criminal Attractiveness	13
3.4 Dataset Overview	13
4 Methodology	15
4.1 Research Approach	15
4.2 Latent Profile Analysis	16
4.2.1 Indicator Selection	16
4.2.2 Model Estimation Procedure	17
4.3 Post-Hoc Analysis	17
4.3.1 Network Analysis: Address-Sharing Patterns	18
4.3.2 Chainalysis: Illicit Exposure Assessment	19
4.3.3 Conditional Profile Distributions for Attribution Mitigation Behaviors	20
4.4 Interviews with Law Enforcement	20
4.4.1 Focus Group Structure and Analysis	20
4.5 Research Flow Diagram	21
4.6 Ethical Considerations	22

5	Conceptual Lens for Behavioral Indicators	23
5.1	Behavioral Domains of Cryptocurrency Exchange Usage	23
5.2	Towards a Conceptual Framework	24
5.2.1	Operational Transactional Practices: Behavioral Dimensions	24
5.2.2	Conceptual Framework	25
5.3	LPA: Data and Selection Criteria	25
6	Empirical Assessment of Behavioral Indicators	27
6.1	Dataset Overview and Scope	27
6.2	Population Structure and Coverage	28
6.2.1	General User Population	28
6.2.2	Financially Active Population	28
6.2.3	LPA Sample Construction	29
6.2.4	Presence of Internal Administrative Accounts	30
6.3	Activity Concentration	31
6.4	Conclusion and Implications for LPA	31
7	Latent Profile Analysis on User Types	33
7.1	Indicator Selection and Data Preparation	33
7.1.1	Operationalization	33
7.1.2	Handling Missing Values and Structural Zeros	36
7.1.3	Indicator Distributions and Transformations	36
7.1.4	Correlations	37
7.1.5	Final Conceptual Model	37
7.2	Results	39
7.2.1	Relaxation of the Local Independence Assumption	39
7.2.2	Model Fit Comparison	40
7.3	Profile Overview	41
7.3.1	Profile 1: Moderately Active Diversified Users	42
7.3.2	Profile 2: Extreme Passive accounts	42
7.3.3	Profile 3: Low-Activity Holders	42
7.3.4	Profile 4: High-Volume Active Traders	42
7.3.5	Profile 5: Systematic Accumulators	43
7.4	Sensitivity Analysis of the Profile Solution	43
7.5	Interpretation Summary and Implications	45
8	Post-Hoc Profile Characterization	47
8.1	Analytical approach	47
8.2	Profiles and Address Sharing Patterns	48
8.2.1	Interpretation and Limitations of the Withdrawal Network	49
8.2.2	Network Metrics for Analysis	50
8.2.3	Withdrawal Network: Results	51
8.3	Profiles and Receiving Exposure	53
8.4	Profiles and Attribution Mitigation	54
8.4.1	KYC-Compliance	54
8.4.2	Use of Privacy Coins	55
8.5	Conclusion	56
8.5.1	Consistency matrix	56
8.5.2	Impact of Internal Users	58
9	Qualitative Insights from Law Enforcement	59
9.1	Challenges in Interpreting and Validating Latent Profiles	59
9.2	Dependence on Contextual Variables	60
9.3	Practical Relevance	60
9.4	Implications for Investigative Research and Detection Bias	60
9.5	Critical Reflections and Limitations	61
9.6	Conclusion	61

10 Discussion	63
10.1 Key Findings	63
10.1.1 Interpretation of the Profiles and Post-Hoc Analysis	63
10.1.2 Integration with Findings from the Focus Group	64
10.2 Methodological Reflection and Limitations	65
10.3 Contributions	66
10.3.1 Methodological Contributions	66
10.3.2 Scientific Contributions	67
10.3.3 Implications for Investigative Practice	67
10.4 Recommendations for Future Research	68
11 Conclusion	69
AI Statement	71
References	72
A Literature review search strategy	77
A.1 Literature synthesis	78
B LPA: Detailed Results	81
B.1 Correlations of Indicator Variables	81
B.2 Relaxation of the local independence assumption	82
B.3 Log-Likelihood Stability Diagnostics	83
B.4 Profile Solutions for the 3-, 4- and 5-Cluster Models	83
B.5 Distribution of Removed Indicator Variables	85
B.6 Sensitivity Analysis	85
B.6.1 Outlier Removal	85
B.6.2 Internal User Account Removal	87
C Supplements for Profile Comparison	89
C.1 Distributions of Indicators across Profiles	89
C.2 Characteristics of Internal Users	89
D Post-Hoc Analysis Supplements	91
D.1 Address-Sharing Network Analysis	91
D.2 Blockchain Coverage and Illicit Exposure	93
D.3 KYC Applications and Verifications	93
D.4 Privacy Coin Use	94
E Focus Group Protocol	95
E.1 Pre-Focus Group Questionnaire	95
E.2 Focus Group Structure	95
E.2.1 Respondents' Introduction	96
E.2.2 Focus Group Protocol	96
E.2.3 Respondent Overview	97
F Focus Group Summary of Results	98
F.1 Current Practices (Q1)	98
F.2 Interpretation and Validation of the Profiles (Q2/Q3)	99
F.3 Relevance and Applicability in Other Contexts (Q3/Q4/Q6)	99
F.4 Limitations and Missing Elements (Q6)	100

List of Figures

2.1	UTXO model by SEBA Bank (2020)	7
2.2	Depiction of sending and receiving exposure within Chainalysis	11
4.1	Research Flow Diagram	22
5.1	Conceptual framework of behavioral domains and the identification and their relation to the latent behavioral profiles	25
6.1	User coverage across transactional tables	29
6.2	Number of deposit and withdrawal transactions per user on a log scale	30
6.3	Transactions per user (internal users) on a log scale	31
6.4	Account creation frequency for the study population	31
7.1	Conceptual model for the LPA	38
7.2	Distribution of indicator variables (log-transformed)	39
7.3	Statistical fit indices for the base model	40
7.4	Mean z-scores for each cluster of the 5-profile solution	41
7.5	Comparison of model fit indices across latent class solutions: Base model and sensitivity analysis	44
7.6	Comparison of model fit indices across latent class solutions: sensitivity analysis without internal users	45
8.1	Proportion of users present in the withdrawal-based address-sharing network across profiles	51
8.2	Distribution of centrality measures for Profiles 1 and 4	52
8.3	User distribution and illicit exposure on deposit addresses per profile	53
8.4	Over- and underrepresentation in blockchain coverage and illicit exposure across latent profiles	54
8.5	Over- and underrepresentation of latest KYC statuses across profiles	55
8.6	Over- and underrepresentation of privacy coins usage across profiles	56
8.7	Comparative patterns of post-hoc analysis across profiles	57
A.1	Literature review flowchart PRISMA (2020)	78
B.1	Correlations between potential indicator variables	81
B.2	Correlations between the final selection of indicator variables	82
B.3	Mean z-scores for removed indicator variables across the 5-profile solution	85
B.4	Comparison of standardized indicator means between the original and outlier sensitivity analysis of the 5-profile solutions	86
B.5	Comparison of standardized indicator means between the original and internal user sensitivity analysis of the 5-profile solutions	88
C.1	Boxplots of LPA indicators across the five latent profiles	89
D.1	Visual representation of withdrawal network based on address-sharing, without internal users	92

List of Tables

3.1	Overview of included tables and their recorded data in the available dataset	14
4.1	Overview of the subquestions their supporting research methods and intended contributions	16
5.1	Operational behavioral dimensions within exchange usage	25
6.1	Overview of the user population	28
6.2	Descriptive statistics per transactional table	29
6.3	Descriptive statistics of transactions per user for deposits and withdrawals (internal users)	30
7.1	Operationalization of Conceptual Behavioral Dimensions	34
7.2	Descriptive Statistics of Indicator Variables	38
7.3	Fit indices and stability diagnostics for latent profile solutions	40
7.4	Median indicator values per latent profile (original scale)	42
7.5	Observations identified as outliers per indicator variable	43
8.1	Classification certainty by latent profile	48
8.2	Network overview per constructed network	49
A.1	Overview of included publications	79
A.2	Overview of included literature and analytical dimensions.	80
B.1	Model fit indices for LPA solutions - without relaxation of the local independence assumption	83
B.2	Log-likelihood values across random start configurations	83
B.3	Standardized indicator means for the 3-profile solution	84
B.4	Standardized indicator means for the 4-profile solution	84
B.5	Standardized indicator means for the 5-profile solution	84
B.6	Model fit statistics — outlier removal sensitivity analysis (500 random starts)	85
B.7	Standardized indicator means for the 5-profile solution — outlier removal sensitivity analysis (500 random starts)	86
B.8	Model fit statistics — internal user account removal sensitivity analysis (500 random starts)	87
B.9	Standardized indicator means for the 5-profile solution — internal user account removal sensitivity analysis (500 random starts)	87
C.1	Distribution of selected users across latent profiles	90
D.1	Network centrality measures and component presence per profile	91
D.2	Legend for profile colors used in network visualizations (corresponding with LPA profiles)	92
D.3	Coverage and illicit exposure deviations across latent profiles	93
D.4	Internal compliance characteristics per latent profile	93
D.5	KYC approval and missing-status deviations across latent profiles	94
D.6	Privacy coin use deviations across latent profiles	94
E.1	Alignment between research objectives and focus group questions	96
E.2	Overview of focus group respondents	97

List of Abbreviations

Abbreviation	Definition
AIC	Akaike Information Criterion
AML	Anti-Money Laundering
AOT	General Investigative Team
BIC	Bayesian Information Criterion
BVR	Bivariate Residual
CDD	Customer Due-Diligence
CEX	Centralized Exchange
CoSEM	Complex Systems Engineering & Management
DEX	Decentralized Exchange
FACT	Financial Advanced Cyber Team
FIOD	Dutch Financial Information and Prosecution service
FIU	Dutch Financial Intelligence Unit
IQR	Interquartile Range
KYC	Know Your Customer
LCA	Latent Class Analysis
LEA	Law Enforcement Agency
LPA	Latent Profile Analysis
ML	Machine Learning
P2P	Peer-To-Peer
UTXO	Unspent Transaction Outputs
VABA	Virtual Asset Business Act
VASP	Virtual Asset Service Provider

1

Introduction

Cryptocurrency exchanges have become central infrastructures within the digital financial ecosystem. By enabling the conversion between fiat currencies to virtual currencies and vice versa (Kruisbergen et al., 2019), these platforms facilitate a broad range of economic transactions with digital finance and foster innovation in digital finance. At the same time, these decentralized, critical interfaces between blockchain-based transactions and the traditional financial system, allow financial flows to move rapidly across jurisdictions and operational environments. As a result, cryptocurrency exchanges have increasingly become facilitators for illicit transactions, in organizing, expanding and laundering illegal financial flows (Schrama et al., 2022).

The growing integration of cryptocurrencies into criminal revenue models has substantially increased the complexity of financial investigations (Chainalysis, 2025). Cryptocurrencies are valued within criminal environments for their pseudo-anonymity, transaction speed, and ability to easily cross borders without direct reliance on conventional banking infrastructures (Bryant & Kennedy, 2014; Shcherbak, 2014). Consequently, cryptocurrency exchanges have become increasingly relevant within money laundering processes. Here, they may be used to obscure the origin of illicit funds and support the reintroduction of criminal proceeds into the legitimate economy (Europol, 2024; Moser et al., 2013). Recent estimates indicate that illicit cryptocurrency-related transaction volumes reached approximately \$51 billion globally in 2024, reflecting almost a 25% yearly increase compared to 2020 (Chainalysis, 2025). Beyond financial crime alone, the misuse of digital financial infrastructures is associated with broader societal risks, including ransomware facilitation, sanction evasion, terrorist financing, and the undermining of financial integrity (Bank for International Settlements, 2022; Europol, 2024). Cryptocurrency exchanges continue to account for a large share of these illicit proceeds (Chainalysis, 2025).

Consequently, cryptocurrency exchanges represent particularly relevant environments to examine how they are used by different users and the variation in their behaviors, due to the facilitating role they have in both legitimate and illicit financial activities. However, such data is not always available. When it is, cryptocurrency exchanges generate substantial volumes of transactional data. This complicates the extraction and interpretation of meaningful behavioral structures within such environments. Transactional behavior within exchange environments is often heterogeneous, involving different operational patterns, transaction activity and structures, and cash-out strategies between users. As these environments and money laundering strategies continue to evolve, distinguishing meaningful behavioral structures and users in large-scale transactional data becomes increasingly challenging. This is particularly relevant in contexts where illicit actors may deliberately attempt through tactics to blend into broader transaction flows rather than exhibit clearly anomalous behavior (Fu et al., 2025). Consequently, understanding how different behavioral configurations manifest within exchange environments remains an important challenge for both research and practice.

This challenge is particularly relevant in law enforcement contexts. Due to increasing volumes of digital financial data and limited investigative capacity, law enforcement agencies (LEAs) are increasingly required to prioritize investigative efforts and interpret complex transactional environments efficiently

(Dutch Ministry of Justice and Safety, 2024). This development aligns with a broader shift toward intelligence-driven investigative practices, in which analytical interpretation and prioritization play an increasingly central role. In such contexts, understanding behavioral patterns and operational structures may be more valuable than solely identifying isolated suspicious transactions. Moreover, as cryptocurrency exchanges fulfill a central position within digital financial ecosystems, they represent strategic points of intervention for disrupting criminal infrastructures (Whelan et al., 2025). Within cyber-enabled crime environments, disrupting facilitating infrastructures may generate broader impact than focusing exclusively on individual offenders, as it directly affects the scalability and operational continuity of criminal activity.

1.1. Formal Takedown of Cryptex

Seized cryptocurrency exchanges provide rare opportunities to examine actual user behavior within financial infrastructures associated with illicit activity. The takedown of virtual cryptocurrency exchange Cryptex, which was involved in large-scale money laundering, is a successful example of a law enforcement intervention aimed at disrupting digital financial infrastructures. Having processed over \$5.88 billion worth of crypto transactions while active, the platform provided financial services to cybercriminals in addition to its mainstream services (Chainalysis, 2024). Cryptex was specifically targeted for laundering over \$51.2 million originating from ransomware operations and processing \$720 million in transactions linked to various cybercriminal enterprises (TRM Labs, 2024).

In September 2024, Cryptex was taken down as part of *Operation Endgame*. The Dutch Financial Information and Investigation Service (FIOD), together with other international partners, succeeded in seizing €7 million in criminal funds (Belastingdienst, 2024). This intervention has resulted in a unique dataset with detailed information on user characteristics and transactional activity on the platform. Since such databases are usually inaccessible to researchers, this provides an opportunity to derive rare and valuable insights.

Analyzing user activity and associated transactional behaviors within this dataset enables LEAs to better understand the (criminal) use of such cryptocurrency exchanges and patterns of activity performed on the platform. These insights are considered to inform and support the prioritization and refinement of enforcement strategies for safeguarding the integrity of the decentralized financial system.

1.2. Previous Research and Knowledge Gaps

The previous sections highlighted the societal relevance of this research. This section discusses the academic relevance, based on 15 studies that form the initial body of knowledge in this field. From this review, three knowledge gaps are identified that motivate the problem statement and objective of this research. The search strategy as well as an overview of the 15 key included studies for this literature review can be found in Appendix A.

1.2.1. User profiles and Behavioral Aspects Related to Cryptocurrency Exchanges

Existing literature prominently applies machine learning (ML) for classifying to which exchange service a specific wallet belongs (Kim et al., 2025; Pettersson Ruiz & Angelis, 2022), or they propose ML frameworks to build customer detection systems for these exchanges (Jiang et al., 2022; F. Zhou et al., 2024). While this provides valuable insights into transaction flows, these frameworks and models remain black boxes that offer limited understanding of the behavioral mechanisms and trade-offs that play a role in the underlying financial crimes. This observation is underscored by Fu et al. (2025), who mention in their Systemization of Knowledge (SoK) paper that analyzing money laundering behaviors not only requires innovative methods and technologies, but also a more profound understanding of underlying patterns. Hu et al. (2024) focus on the scale and mechanics of instant cryptocurrency exchanges and show how some patterns can be detected, but link this to aggregate behavior, instead of specific illicit user profiles. More recently, Lee et al. (2025) analyze the behavioral patterns of centralized cryptocurrency exchanges by examining transaction-level features such as transaction fees, transfer periods between deposit and hot wallets, and the number of transfers per transaction. While this study shows that meaningful behavioral distinctions can be drawn from transactional data, this study operates at the exchange-level as an entity, rather than the individual users. Therefore, the behavioral mechanisms of individual users remain unexplored. In addition, Zola et al. (2025) study the behavioral responses

after law enforcement actions, through a flow and network analysis. Even though they provide important insights through behavioral patterns in blockchain data, the research focuses on the blockchain, instead of internal cryptocurrency exchange data and does not define different user profiles that could explain the observed behavioral mechanisms. Other studies, such as Tovanich et al. (2025) mention the potential for user type classification.

Hu et al. (2025) and Gu et al. (2022) both focus on analyzing anomalies in transaction amounts on cryptocurrency exchanges and the prediction of such occurrences, yet they do not explore the behavioral aspects that can explain criminal *modus operandi* in greater depth. Reynolds and Irwin (2017) find that although law enforcement can theoretically expose illicit users of exchanges, weak compliance practices and verification control allow them to remain anonymous, which highlights the importance of behavioral analysis, above an evaluation of the compliance and verification procedures. Lastly, although some criminological work studies the presence of concrete user profiles, such as Booij et al. (2021), van de Laarschot and van Wegberg (2021), and Van Wegberg et al. (2020), a perspective like this has not yet been applied to cryptocurrency exchange data with a focus on transactional patterns.

1.2.2. Access to Empirical Exchange-Related Data

Most of the included studies in this review use open blockchain or scraped data, which leads to system-level analyses rather than user-level insights specifically related to an (illicit) exchange service. This is also underscored by Fu et al. (2025). Internal exchange datasets remain largely inaccessible (Kim et al., 2025), which limits opportunities to study extensive behavioral patterns of users. F. Zhou et al. (2024) and Pettersson Ruiz and Angelis (2022) form the only exception in the selected literature, as F. Zhou et al. (2024) collaborate with actual exchange services to analyze their internal data and Pettersson Ruiz and Angelis (2022) use data made available by a LEA to develop an algorithm contributing to exchanges' AML-efforts. However, none of the prior studies investigate seized exchanges already known to be tied to criminal activity, nor do they apply a strong behavioral lens. Moreover, their findings were not directly aimed at improving the efforts of law enforcement agencies (LEAs). Fu et al. (2025) highlight the limitation of existing studies that transaction datasets often include only basic information and lack deeper context. This makes it harder to fully understand the behavior of the transactions. They argue that future research could be improved by incorporating a wider range of transaction types and more context-related information.

1.2.3. Limited Inclusion of Law Enforcement Perspectives

Only a few studies explicitly consider engaging or considering the requirements of law enforcement practice. Much of the literature briefly touches upon practical implications (van Wegberg et al., 2018; F. Zhou et al., 2024), but without direct engagement of LEAs. Pettersson Ruiz and Angelis (2022) include interviews to reflect on their findings, but not with law enforcement experts. As mentioned, many studies aim to provide tooling aiding AML-practices, but these are mainly framed from the perspective of exchanges and their current KYC-procedures (Jiang et al., 2022; F. Zhou et al., 2024). Similarly, while discussing the difficulties of deploying successful sanctions by LEAs, Zola et al. (2025) remain rather analytical and do not directly involve investigative authorities in their research. As such, much of the existing work focuses on applications for exchange services, only superficially considering law enforcement relevance. This creates a gap for properly combining academic insights into investigative practice. However, this can be considered crucial as ML-models and automated frameworks cannot always be directly applied in policing contexts due to the requirements of transparency and good governance. This again highlights the strong need for *understanding* criminal behavioral aspects and trade-offs that are missing.

1.3. Thesis Scope

This section defines the scope of this thesis, by discussing a summarizing problem statement, the associated research objective and following research questions. Moreover, this section argues the link to the Complex Systems Engineering and Management (CoSEM) study programme.

1.3.1. Problem Statement

Previous sections highlighted the importance of cryptocurrency exchanges in facilitating illicit financial behavior and consequently, the need for understanding their use from a law enforcement perspective.

Existing studies mainly focus on detection frameworks for illicit practices on cryptocurrency exchanges, often relying on public data. However, previous academic work has not examined (criminal) user profiles and their associated transactional behaviors related to an empirical exchange dataset, nor has it assessed how these findings translate into implications for law enforcement practice. Together, these gaps highlight limited behavioral understanding grounded in empirical data and operational practice.

1.3.2. Research Objective

Based on the scientific and societal relevance that emerge from the introduction and identified knowledge gaps, the following research objective is formulated:

To identify and interpret latent user profiles and transactional structures within internal data from a seized cryptocurrency exchange, and to examine how these behavioral patterns can support risk-based investigative strategies in law enforcement practice.

1.3.3. Research Questions

Together, subsection 1.3.1 and subsection 1.3.2 lead to the following main research question:

To what extent can latent behavioral profiles be identified within the internal records of a seized cryptocurrency exchange, and what insights do these profiles provide for law enforcement investigations?

Given that this research is particularly based on the Cryptex database as outlined in section 1.1, the following sub-questions are formulated to address the main research question:

1. ***SQ1: What types of user behavior can be identified as potentially meaningful for distinguishing behavior within crypto-exchange environments?***

This sub-question focuses on identifying which characteristics of exchange use are relevant indicators of money laundering by conducting a literature review. Indicators such as activity levels, transaction patterns or account attributes are potentially informative for distinguishing between user patterns. This step is essential to the analysis and serves as a preparation for SQ2.

2. ***SQ2: What latent user profiles can be identified within the Cryptex exchange?***

This sub-question aims to identify different user types in the Cryptex database by means of a Latent Profile Analysis (LPA). By including relevant user indicators derived from SQ1, users can be classified into meaningful profiles. This contributes to a better understanding of the typology of actors active on the exchange and provides a basis to explore whether specific transactional behaviors and criminal modus operandi can be linked to these profiles.

3. ***SQ3: How do the identified user profiles differ in behavioral patterns associated with increased risk?***

Building on the previous sub-question, this part contextualizes the identified profiles by two complementary analyses. First, internal classifications and risk labels assigned by Cryptex are examined to assess whether the behavioral differences of the profiles are reflected in the exchange's own risk assessment. Second blockchain data is analyzed to uncover how the identified user profiles operate within the broader financial network. Together, these analyses aim to enrich the behavioral interpretation of the profiles, to better assess their relevance for law enforcement.

4. ***SQ4: What insights can be derived from the identified profiles and their associated risks patterns?***

This sub-question aims to interpret the enriched profiles of the previous research question by means of a literature review and expert interviews, to examine if certain user profiles exhibit behavioral and structural patterns associated with criminal activity that can be potentially relevant for identifying illicit behavior.

5. ***SQ5: How can the obtained results inform and support investigative strategies of law enforcement agencies in practice?***

Lastly, through expert interviews, this sub-question evaluates how the findings can be translated into practical value for law enforcement. For example, findings could indicate behavioral indicators, objective risk profiles, or approaches that may support or improve intelligence-led investigations.

1.3.4. Link to the MSc CoSEM Programme

This research is conducted as part of the MSc programme in Complex Systems Engineering and Management (CoSEM) at Delft University of Technology, which integrates a multi-disciplinary perspective including the analysis of technical, social, regulatory, and institutional aspects of various socio-technical systems to consider better design and governance. This research provides a technical analysis of a cryptocurrency exchange using case data to examine patterns in criminal utilization, whilst taking into account the broader social and institutional setting. The objective is to formulate recommendations for LEAs in supporting their investigative operations. Aligned with the CoSEM programme, this study combines technical analysis, systems thinking, and institutional perspectives to understand and improve the governance of digital financial infrastructures.

1.4. Thesis Outline

This document is structured as follows. Chapter 2 provides the technical and contextual background on virtual currencies, blockchain technology, cryptocurrency exchanges, and their relation to illicit financial practices. Chapter 3 contextualizes the research by discussing the case of Cryptex in more detail. Chapter 4 describes the methodology and research steps undertaken to address the research questions. Chapter 5 develops the conceptual framework and data requirements guiding the selection of indicator variables for the LPA. This is empirically examined using the dataset described in Chapter 6. Chapter 7 presents the results of the LPA and the identified user profiles. Chapter 8 presents the post-hoc analysis, aimed to further enrich the interpretation of these profiles. Chapter 9 reports the findings from qualitative research that focuses on translating the results to law enforcement practice. Finally, a discussion and conclusion is provided respectively in Chapter 10 and Chapter 11 of this thesis.

Technical and Contextual Background

This chapter forms the theoretical and analytical background of this research, defining the core concepts and technical landscape of blockchain-based cryptocurrencies, criminal behavior, as well as the functioning and role of cryptocurrency exchanges, which guide the reader to the subsequent chapters of this research. Here, we use *Bitcoin* as a reference blockchain to illustrate relevant mechanisms, as this was one of the primary traded, crypto-assets within Cryptex. Importantly to mention, other blockchain assets were also present, including for example *Ethereum* which uses an *account-based* transaction model. However, the main aim of this chapter is to provide an illustration of the general principles underlying blockchain transactions.

For this research, we do not examine the broader economic implications caused by the presence of Bitcoin, nor delve into technical cryptographic details such as block creation (i.e., mining). Instead, the focus is put on conceptual aspects that are relevant for understanding user behavior on cryptocurrency exchanges.

2.1. Fundamentals of Bitcoin and Blockchain Technology

In 2009, a group of anonymous developers launched the first digital currency functioning on a blockchain: Bitcoin (Nakamoto, 2009). This development significantly altered the perspectives on traditional monetary systems. Bitcoin marked the first ever fully digital peer-to-peer network enabling financial transactions without the need for a central authority (Antonopoulos, 2017). The core innovation of Nakamoto (2009) lies in the fact that the system is completely decentralized: no trust in third parties is required, and transactions are publicly recorded on an unchangeable log, called blockchain. This implies that every single Bitcoin, in principle, can be traced back to its origin through the chain of transactions.

Within the blockchain, anyone can create a Bitcoin account. A Bitcoin address, used to receive funds, is derived from a so-called 'public key', designed to be shared and is similar to a bank account number. It is important to note that a wallet can generate multiple public keys and corresponding addresses. This can in principle be done without any administration procedures, identity requirements, or fees (Böhme et al., 2015). A public key is again derived from a private key, which is a required signature to authorize transactions and must be kept secret by that user, as it can be compared to a PIN-code (Antonopoulos, 2017). The system that manages an entity its public and private keys is called a wallet. This may take the form of a hardware device, software, or even a paper wallet that depicts an entity its public and private key.

Bitcoins are recorded as transactions, and are in fact not "held" by a user (Böhme et al., 2015). When user A wants to transfer money to user B, they send a transaction secured using the public key of user B, and it can only be decrypted by their private key. The blockchain enables the verification that user A has, in fact, the desired value of Bitcoins available. Due to this functioning, there are no built-in mechanisms to undo transaction errors, making them inherently irreversible (Böhme et al., 2015).

As mentioned, a wallet can consist of multiple addresses, which allows users to separate transactions (Nakamoto, 2009). When initiating a transaction, Bitcoin does not track a single account balance. In-

stead, it combines one or more so-called Unspent Transaction Output (UTXO's) to reach the desired transaction amount (Conti et al., 2018). Then, a transaction produces two outputs: one consists of the amount sent to the recipient. The other output is sent towards the change address, which represents the remaining value returned to the sender after spending the selected UTXOs. This mechanism describes what is commonly referred to as the multiple-input spending heuristic (Nakamoto, 2009).

However, this mechanism also makes it possible to derive the relationships between different addresses. Because multiple UTXOs from different addresses may be combined as inputs in a single transaction, we can apply clustering heuristics to examine whether these addresses are likely controlled by the same user. Figure 2.1 below illustrates the mechanism behind the UTXO model, as illustrated by SEBA Bank (2020).

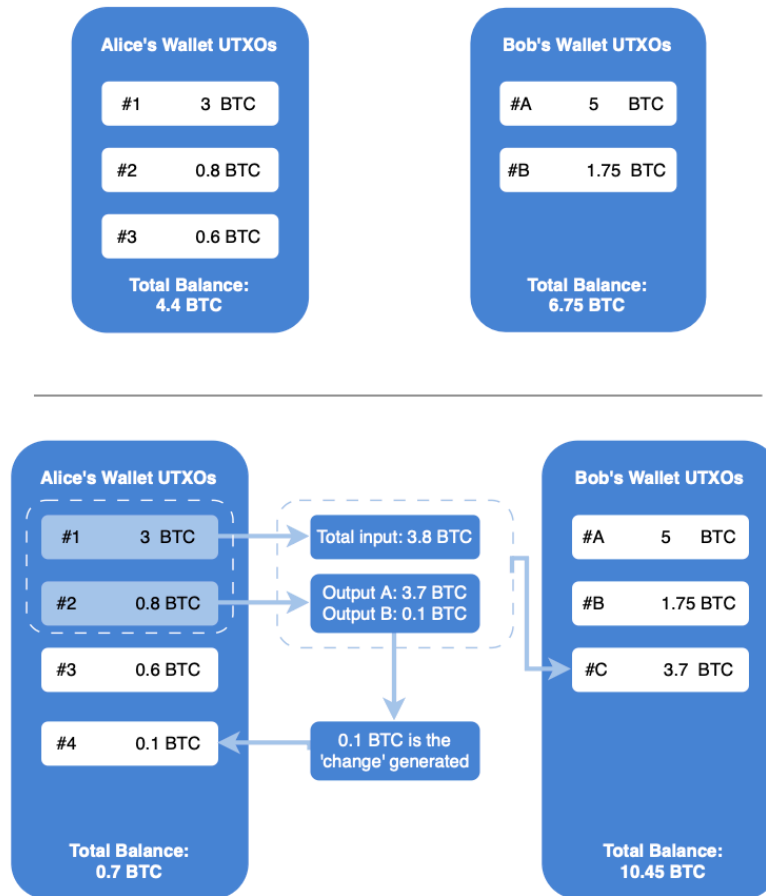


Figure 2.1: UTXO model by SEBA Bank (2020)

While Bitcoin serves as the main reference blockchain in this chapter, it is important to note that Cryptex functioned as a multi-asset exchange that supported several cryptocurrencies and digital assets, of which *USD-stablecoins* and Bitcoin were the most dominantly used in transactions on the platform, followed by *Litecoin*, Ethereum and *Monero*¹. Bitcoin and Litecoin for example, rely on the UTXO-based model as depicted in Figure 2.1, whereas other cryptocurrencies may use other models. Moreover, coins such as Monero and Zcash are so-called 'privacy coins', that enhance privacy features and limit traceability. These differences affect the interpretation of on-chain transaction structures and how addresses can be reliably clustered. The conceptualization in this section however, still provides a helpful framework for understanding transaction behavior related to cryptocurrency exchanges while keeping in mind differences that exist which can influence further analyses within this research.

¹As found in the empirical dataset of this research, further described in Chapter 6.

2.2. Illicit Financial Flows and Money Laundering

Consequently, cryptocurrencies are considered attractive for illicit practices, given their perceived anonymity and transaction speed, as no third-party intervention is required (Shcherbak, 2014). Additionally, the ease with which anyone can create an account, together with regulatory oversight still being under development, creates further opportunities for cybercriminals (Böhme et al., 2015). However, this anonymity is logically only pseudonymous: as all transactions are visible on the public blockchain, once an address is linked to a user, their entire transaction history can be traced (Conti et al., 2018; Moser et al., 2013).

Nonetheless, the pseudonymous nature of Bitcoin has not discouraged criminals from using it for illicit practices. In fact, as described in Chapter 1, criminal adoption is only expected to increase. Böhme et al. (2015) distinguish three main criminal concerns related to Bitcoin. First, *Bitcoin-specific crimes* are recognized, which include attacks on the currency and its infrastructure, or Bitcoin theft. Second, *Bitcoin-facilitated crimes* include the payment by means of Bitcoin for illicit goods or services that can be used offline (e.g. drugs, weapons). Third, and particularly relevant for this research, is the use of cryptocurrencies for *money laundering*. This is seen as a primary concern related to Bitcoin (Europol, 2022). In this study, money laundering is defined according to the formulation adopted by the Financial Intelligence Unit (FIU), based on Article 420bis of the Dutch Criminal Code: “Carrying out (or having others carry out) acts that provide apparently lawful origins for an increase in assets that have been kept from the legal authorities. The purpose of money laundering is to obscure the provenance of money.” (Dutch Financial Intelligence Unit, n.d.).

Typically, the process of money laundering includes three stages: *placement* into the financial system, *layering* to obscure the origin and finally, *integration* into the legal economy (Levi & Soudijn, 2020). Their proceeds often start as fiat currency, such as euros or dollars, which are then converted to enhance anonymization and layering as part of their cash-out strategy (van Wegberg et al., 2018). This is where the relevance of cryptocurrency exchanges arises, which leads us to the following section 2.3.

2.3. Cryptocurrency Exchanges as Criminal Facilitators

While the technical characteristics of section 2.1 illustrate how users manage their transactions on the blockchain, most cryptocurrency activity does not take place solely in this manner. In practice, most users engage with cryptocurrencies through platforms, of which cryptocurrency exchanges are one of the most central intermediaries. These exchanges enable the conversion between cryptocurrencies such as Bitcoin or other virtual assets and ‘regular’ fiat currencies (e.g., dollars or euros) (Kruisbergen et al., 2019). They emerged to simplify the purchase of cryptocurrencies, offering a more user-friendly alternative compared to mining them through specialized software or obtain them through peer-to-peer transactions (Biscontin, 2024). Moreover, to avoid the risk of having to individually manage and protect your public and private keys, lost or stolen devices, specific “custodial” exchanges provide the opportunity to hold your wallets (Conti et al., 2018). This makes them function as an important and widely-used intermediary within the crypto ecosystem, accounting for 90% of all funds sent by services (Chainalysis, 2020; Volten et al., 2025).

It is important to recognize that within the broader crypto-ecosystem, other services exist that can facilitate money laundering schemes and strategies. For instance, mixing services are designed in such way that transactional links on the blockchain are further obscured by pooling and redistributing funds, which complicates any blockchain analysis. Although these services fall outside the scope of this study, acknowledging their role within laundering typologies remains essential. According to research by Fanusie and Robinson (2018), Bitcoin exchanges ultimately received the greatest total amount of illicit crypto, even though they in general process the majority of overall Bitcoin transactions.

2.3.1. Illicit Potential across Cryptocurrency Exchanges Types

Unsurprisingly, a strong relationship exists between these cryptocurrency exchanges and money laundering. Across all stages of the money laundering process as described by Levi and Soudijn (2020) and discussed in section 2.2, exchanges create opportunities for the placement, layering and integration of illicit funds. To do so, criminals need effective strategies that support these processes (Levi, 2015). As a result, they have a strong facilitative function in financial crimes conducted with cryptocurrencies (Volten et al., 2025). According to research by Fanusie and Robinson (2018), as exchanges process

the most Bitcoin transactions of all conversion services, they have also processed greatest absolute amount of illicit cryptocurrencies. Moreover, previous research by Leuprecht et al. (2022) indicates that cryptocurrency exchanges appear frequently in datasets related to illicit financial activity. After transfers through traditional banking systems, converting funds between fiat and cryptocurrencies through such third-party exchanges is one of the most common ways to move illicit funds.

Cryptocurrency exchanges can be classified into different types. First, a distinction can be made between centralized exchanges (CEXs) and decentralized exchanges (DEXs), following the definitions of Hägele (2024). CEXs are centralized cryptocurrency trading platforms that offer buying and selling of digital assets in a user-friendly environment. Generally, CEXs operate under federal and state regulations, which enhances their authenticity and security. At the same time, limited transparency at certain exchanges raises concerns about potential illicit practices (Alexander et al., 2023). Because these CEXs operate with an off-chain order-book system, this can be executed much faster than on-chain transactions, which is applicable for most "Automated Market Maker" DEXs. Only deposits and withdrawals are settled on the blockchain, whereas the trading process itself occurs off-chain (Aigner & Dhaliwal, 2021). On the other hand, DEXs function without a central authority and facilitate peer-to-peer cryptocurrency trading by using smart contracts for transactions. This allows users to maintain more control over their digital assets (Chainalysis, 2019b).

Moreover, exchanges can be custodial or non-custodial. As mentioned, custodial exchanges have control over your assets, as they hold the associated private keys of the wallets (Chainalysis, 2020). Big, centralized exchanges tend to be custodial, as it provides convenience for the user. However, trustworthiness of the platform is essential as it also provides custodial risks, of which a good example is the collapse of cryptocurrency exchange FTX (Conlon et al., 2023). Most of the executed transactions within these exchanges happen off-chain. This implies that it is not recorded on the blockchain but instead, are handled by the exchange itself. This enhances speed, but reduces transparency (Chainalysis, 2020).

In contrast, non-custodial exchanges allow users to maintain full control over their private keys and are often decentralized or peer-to-peer (P2P). Usually, no account is required, with no or minimal Know-Your-Customer (KYC) requirements, which increases anonymity, but inevitably poses increased risks for illicit activity (AMLC, 2022).

Within the non-custodial exchange types, instant exchanges form an important subgroup. These platforms bring together many other exchanges in one centrally operated interface, which matches buyers and sellers through an automated process (Chainalysis, 2019b). You receive your asset at a fixed rate as indicated by the platform. Their key advantages include ease of usage, fast execution of transactions and the seamless conversion among multiple fiat and cryptocurrencies. Moreover, there is no public order-book present, contrary to DEXs, resulting in less transparency (Crypto.com, 2024). A desirable feature for cybercriminals, also further explained in 3.3. By repeatedly swapping between assets, criminals can exploit these platforms to layer their illicit proceeds, a technique known as multiple asset conversion.

These (criminal) benefits and functionalities are closely tied to the extent to which exchanges comply with regulatory requirements. Currently, CEXs rely on KYC-principles and Customer Due Diligence (CDD) to identify users in regulated countries, track transactions, and reduce abuse opportunities (Europol, 2022; Ostern & Riedel, 2021). However, the effectiveness of these measures differ greatly between platforms and jurisdictions (Volten et al., 2025). Still, criminals can avoid or manipulate these principles, for example advantage can be taken by exchanges that do not follow strict verification procedures. Or, exchanges themselves take steps or purposefully operate to support cybercriminals, which logically brings us to Chapter 3, outlining the case of Cryptex for this research.

2.4. Common Money Laundering Practices

As discussed in the previous sections, cryptocurrency exchanges and blockchain infrastructures create opportunities for both legitimate financial activity, as well as illegitimate. Within these cryptocurrency ecosystems, criminals are known to deploy various strategies and techniques to obscure the origin and movement of funds, often specifically tied to the three stages of money laundering as identified by Levi and Soudijn (2020). This section provides a global overview of the most prominently adopted strategies.

During the placement stage, illicit funds are typically introduced into the blockchain ecosystem by using multiple accounts and large-value transactions (Fu et al., 2025), as well as exploiting the lack of KYC enforcement (Europol, 2022). However, due to transaction monitoring systems and AML-reporting at regulated exchanges, criminals may also fragment placement transactions into smaller amounts instead of such large-value transactions to avoid visibility, also known as *smurfing* (Zhou, 2025). Subsequently, the layering stage is characterized by more complex transactional behavior aimed at obscuring the origin and movement of funds. Commonly described tactics include fragmented transfers, as well as high-frequency transactions across multiple addresses and rapidly draining wallet balances shortly after receiving funds. Moreover, criminals may employ *chain hopping* strategies by repeatedly converting funds across cryptocurrencies, for example through *swapping* (Zhou, 2025). The use of privacy coins is popular, often through cryptocurrency exchanges or other conversion services such as mixers to complicate tracing efforts (Fu et al., 2025). Finally, during the integration stage funds are often consolidated and transferred to exchange services or other withdrawal providers or intermediaries which support the re-entering of assets into the legitimate financial system. Depending on the specific laundering strategy, this may also involve off-chain cash-out mechanisms, such as cash-based withdrawal strategies (Zhou, 2025).

Particularly the layering stage is considered analytically relevant by Fu et al. (2025), as these strategies may produce observable transactional structures and behavioral patterns within blockchain and exchange data. What becomes clear is that many of these practices revolve around the structuring of transactions, which forms an important aspect of the conceptual lens in Chapter 5.

2.5. Investigative AML Tooling and Blockchain Intelligence

This section provides a general overview for the analytical approaches used to identify and interpret transactional structures and potentially suspicious behavioral patterns within cryptocurrency ecosystems. Rather than presenting an exhaustive overview, this section aims to outline the most prominent approaches discussed in recent literature, summarized by (Fu et al., 2025). Generally, these approaches rely on combinations of heuristic clustering, machine learning (ML) and graph-based analysis or transaction tracing.

2.5.1. Analytical Approaches in Crypto AML

Heuristic rule-based approaches are one of the most commonly applied techniques within crypto AML efforts. These methods rely on predefined transactional patterns and clustering heuristics to identify potentially suspicious activity. For example, through address attribution, exposure analysis or transaction tracing. Because of their relatively high interpretability and practical applicability, this makes them suitable for fast deployment and operational monitoring (Fu et al., 2025). However, since these methods rely on fixed assumptions and rules, they are potentially less effective in adapting to evolving laundering strategies and can produce false positives when legitimate behavior resembles suspicious transactional patterns (Fu et al., 2025).

In addition, ML approaches are used to identify anomalous behavioral structures and transaction features within blockchain ecosystems based on historical data. Lastly, a distinction is made between graph-based methods, which can identify structural relationships, fund flows and transactional clusters that may remain less visible through isolated transaction analysis alone (Fu et al., 2025). Particularly graph-based methods are considered relevant in cryptocurrency investigations due to the inherently networked structure of blockchain transactions (Tovanich et al., 2025). At the same time, as also partly discussed in Chapter 1, limitations exist for both of these approaches regarding interpretability, the evolution of laundering tactics and the difficulty of incorporating broader transactional context into such analyses. Moreover, as criminal behavior adapts to existing mechanisms for monitoring, the question arises whether anomalies in transactions or behavior remain the most relevant indicators for investigation, especially when offenders may intentionally minimize visibility by attempting to blend in with regular activity.

Consequently, these prominent approaches described in recent research emphasize the value of complementing such analyses with broader and more transparent behavioral interpretations of transactional activity that do not solely rely on predefined rules or historical data. Such approaches may provide additional insight into platform-wide behavior. As criminal *modus operandi* continue to evolve and adapt

over time, this underscores the relevance and purpose of this research. Moreover, they may enable the identification of behavioral groupings across users and reveal common transactional practices, which supports the detection of emerging or novel patterns. This is specifically relevant in exchange environments, which function as a facilitating infrastructure, as argued in Chapter 1 and further discussed in section 2.3, as they connect virtual and fiat currencies, thereby creating opportunities for the scaling and concealment of money laundering with cryptocurrencies.

2.5.2. Chainalysis

Chainalysis is a blockchain data platform offering insights into transactional relationships and fund flows, showing similarities to a graph-based approach. As described, blockchain systems are considered pseudo-anonymous. While wallet addresses are not directly linked to real-world entities, all transactions are recorded on the blockchain and publicly visible, creating a network of addresses and transactions (Nakamoto, 2009). Due to models such as UTXO for instance, as used in Bitcoin, it becomes possible to reconstruct those transactions and identify relationships between addresses by means of heuristics and clustering techniques (Conti et al., 2018). This clustering of addresses, complemented with open-source information on entities and forensics, makes it possible to identify and attribute the entity that owns the cluster, together with a label that indicates the type of service it provides (e.g., non-KYC exchange, or gambling) (Chainalysis, 2022).

A core concept within Chainalysis is the notion of exposure, which refers to the degree to which a wallet or entity is connected to certain types of labeled (illicit) activity (Chainalysis, 2023). This can include both direct exposure, where funds are received from or sent to a labeled entity, and indirect exposure, where funds are linked through multiple intermediary transactions, which are called "hops". These exposure metrics are often used as indicators of potential risk and illicit activity, for example in investigations and other anti-money laundering (AML) contexts.

Next to direct and indirect exposure, a distinction is made between receiving and sending exposure. This is clarified in Figure 2.2 below based on (Chainalysis, 2023), where *Wallet 1* has direct receiving exposure from the *Exchange* and *Wallet 1* has sending exposure to *Wallet 2*. This distinction forms an important aspect of the analysis conducted in Chapter 8.

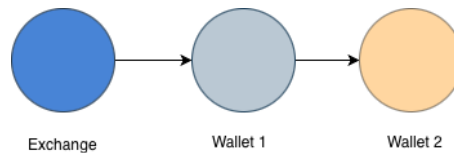


Figure 2.2: Depiction of sending and receiving exposure within Chainalysis

Chainalysis is often part of operations by Dutch law enforcement agencies (Belastingdienst, 2024; Lubbertsen et al., 2025). However, the interpretation of the exposure metrics is not straightforward. The reliance on clustering heuristics introduces uncertainty, as they are based on probabilistic assumptions instead of deterministic identification, although research on classification accuracy proves high certainty (Lubbertsen et al., 2025). However, attribution to a certain cluster does not directly conclude something about intent, as many blockchain services are used by both legitimate and illicit actors, which may lead to ambiguity in the interpretation of interactions with such entities. This is particularly considered relevant for the concept of indirect exposure, where connections are established through multiple transaction hops (Chainalysis, 2023). Similar challenges in interpretation apply to sending exposure. This is because outgoing transactions do not necessarily indicate ownership or control over the destination addresses and may involve initial sending to intermediary services or shared infrastructure.

Despite these limitations, blockchain intelligence tools such as Chainalysis provide a powerful means of structuring and analyzing the origin of large-scale transaction data, as well as attribution of offenders for law enforcement.

Case description: Cryptex

This chapter describes the context and key characteristics of the case that forms the empirical foundation of this research. It provides an overview of the relevant background, operational processes, stakeholders, and eventual takedown of the cryptocurrency exchange. By outlining these aspects, this chapter provides the necessary context and the nature and origin of the data used in this study.

3.1. Investigation by Law Enforcement

In September 2024, the servers of instant cryptocurrency exchange *Cryptex* and *PM2BTC* were seized by the FIOD, as both services were suspected of facilitating illicit financial proceeds and operated infrastructures located in the Netherlands (Belastingdienst, 2024). This research was initiated after initial analysis indicated that Cryptex was involved in various criminal financial flows. Simultaneously, research on PM2BTC showed parallels with the prior investigation. Consequently, the two investigations merged.

This resulted in a formal takedown of Cryptex by Dutch authorities. Moreover, within the US, they were named responsible under American Executive Orders for enabling cybercrime and operating in Russia's financial sector (TRM Labs, 2024). More than \$51 million of ransomware funds were processed, and more than \$720 million had been sent to other high-risk services, such as darknet markets, mixers, and other non-KYC exchanges (TRM Labs, 2024). Moreover, PM2BTC was labeled as a "primary money laundering concern" due to its involvement in Russian criminal finance. Based on findings by FinCEN, it was found that over 50% of PM2BTC's operations were tied to criminal activities, and sophisticated techniques were applied to avoid detection by law enforcement (TRM Labs, 2024). The operation and confiscation of servers used by Cryptex yielded a unique dataset. Within this research, this serves as the primary data source and focus of this study.

3.2. Services Offered

Based on the information obtained through the dataset, as well as online public sources, Cryptex is best considered as a CEX that offered four distinctive services (section 2.3). First, users were able to open a custodial wallet and hold a user balance, which enabled them to deposit, hold and withdraw virtual assets within the platform. Second, Cryptex provided an instant swap service, that allowed users to exchange cryptocurrencies directly without any delay. Third, the platform facilitated specific P2P transfers on the platform, which enabled users to transact directly with other Cryptex users internally without any fees (Business Wire, 2021). Lastly, allowing the conversion between fiat currencies and cryptocurrencies was also possible through cash-drop offs in more than 10 countries (Business Wire, 2021; PR Newswire, 2021). Notably, the platform did not impose mandatory KYC verification for these services, meaning users could access these services with a high degree of anonymity by solely providing an e-mail address, further discussed in the subsequent section.

3.3. Criminal Attractiveness

Cryptocurrency exchange Cryptex is assumed to be considered attractive for criminal practices for several reasons. Here, 6 strongly assumed factors are mentioned, supported by literature that are reflected in the available dataset.

1. **Active marketing on absence of KYC-procedures and limited regulatory authorization:** Criminals tend to use exchanges with weak or absent KYC-procedures to launder illicit proceeds (AMLC, 2022). The same applies to Cryptex, which explicitly advertised “complete anonymity” when registering for an account (United States Attorney’s Office, 2024). Users were only required to register an e-mail and password to create an account (Antoniuk, 2024). This emphasis can be appealing to criminals who desire to obscure their identity as well as transaction history. Moreover, Cryptex’s registration in St. Vincent and the Grenadines, suggests a form of “regulatory arbitrage”, a practice in which firms exploit the differences in different regulatory frameworks to their own advantage (Anti-Money Laundering Network, n.d.). St. Vincent and the Grenadines has historically exhibited limited oversight and supervision jurisdiction on Virtual Asset Service Providers (VASPs) in relation to money laundering risks. Even though the Virtual Asset Business Act (VABA) was introduced in 2022, an operational licensing regime for VASPs has only become effective in 2025 (St. Vincent and the Grenadines Financial Services Authority, 2025). This implies that many crypto platforms were able to operate without much supervision, underscoring the inconsistent and weak enforcement during the period in which Cryptex was active. However, it is considered that such nations remain to have little capacity available for regulatory oversight.
2. **Fast and instant conversion:** Cryptex was an *instant* cryptocurrency exchange that operated as a trading and conversion platform (Chainalysis, 2024). This enabled immediate asset swaps, as described in section 2.3, which is particularly beneficial for criminals seeking to rapidly move their funds with minimal friction. All users were required to do was enter the trade they want to make. Compared to DEXs, instant exchange services generally were considered easier to use (Chainalysis, 2019b), and provide more privacy as transactions are not visible on the blockchain lowering barriers for criminals.
3. **Multiple asset conversion:** As the platform supported the exchange of multiple virtual assets, such as Bitcoin and Ethereum, users were able to quickly move proceeds between different cryptocurrencies and obscure the transaction history even further, making them difficult to trace (Europol, 2022). This method is known as multi-asset swapping and is widely recognized as a common laundering technique among illicit actors.
4. **No public orderbook (instant):** Cryptex operated as an instant exchange, which enabled users to convert assets directly without access to a publicly visible order book (Chainalysis, 2019b). In addition to instant conversions, the platform allowed users to hold balances in custodial wallets and to perform deposits and withdrawals. Contrary to order book-based exchanges, including some DEXs, the Cryptex dataset does not seem to present bid and ask information to users, which may have contributed to a perception of increased anonymity on the platform.
5. **Provision of cash drop-offs:** As described, multiple services were provided, of which one includes cash drop-offs in more than 10 countries (Business Wire, 2021). Presumably, this lowered the barriers for converting offline illicit proceeds into digital currencies (Riccardi & Levi, 2018). Given that the use of cash money is still widely used for money laundering due to its anonymity and limited traceability, such a service such a service would have made Cryptex particularly attractive to criminal actors.

3.4. Dataset Overview

The data used in this research consists of an extensive primary back-office dataset from the cryptocurrency exchange Cryptex. The dataset originates from Cryptex’s back-office and contains detailed system and transaction logs recorded between April 2017 and September 2024, with activity levels varying over time. However, public operations started in 2018 (Chainalysis, 2024) and a website migration took place in March 2023.

The dataset is stored in a relational database that includes about 124 connected tables. Relationships

are established by means of foreign keys that link records to corresponding entries in other tables. Together, they capture information about user accounts, transactions, balances, system events, and operational logs generated by the exchange. Below, in Table 3.1, the key content of this dataset is outlined. Chapter 6 provides a more detailed and exploratory analysis of Cryptex’s back-office operations and user activity, while defining the population and suitability criteria for the subsequent Latent Profile Analysis (LPA).

Analytical aspect	Relevant table(s)	Description of recorded data
Users	auth_user, core_userprofile	Unique user identifiers, account status, registration and profile information
Transactional information	core_transaction, core_order, core_cryptotransaction, core_withdrawal, core_deposit	Transactional actions, either by users or for internal rebalancing
Trading pairs and currencies	core_pair, core_currency	Base and quote currencies, fee structures, and pair characteristics
User balances	core_userbalance	User-level balances across currencies
User activity logs	core_useroperation, core_userhistory	Internal account operations and login-related metadata
Bots and automation	bots_*	Various tables on automated trading, arbitrage, and liquidity management processes
System and communication logs	core_notifications, core_emailmessage	System notifications and user communications

Table 3.1: Overview of included tables and their recorded data in the available dataset

4

Methodology

This chapter describes the design of this study and methods to address the research objective and sub-questions. The overall mixed-method research approach and its argumentation are discussed in section 4.1. The succeeding sections more specifically describe the applied methods to address the sub-questions presented in subsection 1.3.3. Section 4.5 finalizes with a visual representation of the research components and their interrelation within this study. section 4.6 ends with important ethical considerations.

4.1. Research Approach

As this research aims to provide insights into the (potentially criminal) *modus operandi* on a seized cryptocurrency exchange and contrast them with law enforcement practice, a *mixed-method* approach is adopted. This combines a quantitative, statistical analysis to identify distinct user profiles and a complementary contextual data analysis, supported by qualitative interviews that examine the perspectives of and usefulness for LEAs. Together, this enables a more comprehensive understanding of the research problem and potentially aid investigative strategies.

Although various definitions exist, a mixed-method approach generally integrates both quantitative and qualitative components to address the research questions. By doing so, it provides a more holistic understanding of the research problem at hand than a single method could achieve (Johnson et al., 2007). Moreover, methods can validate each other's findings, thereby increasing confidence in the obtained results (Jick, 1979).

Various typologies have been developed to structure how methods can fit into a mixed-method research design (Guest & Fleming, 2015). In this study, the typology provided by Creswell and Clark (2011) is followed, and an *explanatory sequential design* is applied. In this two-phase approach, the research begins with collecting and analyzing quantitative data, which is further examined and explained through qualitative research (Creswell & Clark, 2011). Within the typology of mixed-method research, emphasis is also put on weighting, by giving a specific method priority in the research design. In line with terminology used by Morse and Niehaus (2009), a *QUAN* → *qual* design is adopted. This implies that the qualitative interviews do not represent the central dataset of this study, but instead serve as an important supplementary method to answer the main research question.

By addressing a key limitation of previous research described in section 1.2, namely the lack of interpretability in understanding criminal behavior and translating results onto investigative practice, this mixed-method approach introduces a novel research design in the current academic field on this topic. Given the knowledge gaps indicated in section 1.2 and the absence of research of this kind, this study is considered an exploratory first step, with results that can provide a valuable foundation and direction for future research.

Table 4.1 provides an overview of all research components included in its research approach, along with their intended contribution in relation to the scope of this thesis, as outlined in section 1.3. For each of the methods and their implementation discussed in the remainder of this chapter, methodological

choices were made in such way, that they align with the intended contributions.

Table 4.1: Overview of the subquestions their supporting research methods and intended contributions

Subquestion	Method	Intended Contribution
SQ1	Policy literature review combined with descriptive and exploratory data analysis	Initial understanding of the dataset, supporting the indicator selection and providing context for subsequent analyses.
SQ2	Latent Profile Analysis (LPA)	Identification of distinct user profiles based on transactional behavior, providing a structured representation of heterogeneity within the dataset.
SQ3 & SQ4	Post-hoc analysis (e.g., network / KYC / exposure)	Further characterization and interpretation of the identified profiles by linking them to additional variables and contextual indicators.
SQ4 & SQ5	Focus group with experts	Validation of the representativeness and the practical relevance of the obtained profiles, and insights into their applicability within investigative contexts.

4.2. Latent Profile Analysis

To address SQ2, a Latent Profile Analysis (LPA) will be conducted to identify user types with shared exchange activity and characteristics within the Cryptex dataset. It is assumed that a user's intent, here: its motive, either legitimate or illegitimate, has an influence on the observed behavior. LPA is often used in behavioral research, and therefore serves as a promising method to gain insights into transactional behavior on a cryptocurrency exchange. The underlying assumption of an LPA is unobserved heterogeneity, meaning that it identifies unobserved group structures based on patterns which observed variables display in the data (Spurk et al., 2020; Weller et al., 2020). This is often the case in an investigative context, where a single high value on an indicator may be an interesting signal, but a specific pattern of behavior may be even more indicative (Vermunt, 2004). In this way, the multidimensional behavioral characteristics of users can be captured (Weller et al., 2020).

Unlike earlier studies that relied on opaque machine-learning clustering, LPA improves interpretability because the resulting profiles are more transparent to understand and can capture patterns that may reflect underlying criminal *modus operandi*. This approach may help investigators identify recurring patterns of activity and prioritize high-risk profiles, potentially improving the understanding of facilitator use and investigative efficiency. Models based on such (criminal) profiles are assumed to have a more increased explanatory power (Van Wegberg et al., 2020). Importantly, within LPA, the relationship of these observed indicators and the underlying latent profiles is probabilistic. This means that users are not definitively classified into a single type, but rather have a certain probability (Vermunt, 2004). This probabilistic approach aligns with how uncertainty is usually handled in law enforcement contexts, as it deals with inherent uncertainty of the analysis and supports indicative interpretation, rather than conclusive (Aitken et al., 2010). The main advantages of LPA are therefore transparent and probabilistic grouping, theoretically grounded interpretation, and potential gains in investigative efficiency. Moreover, LPA allows for inclusion of variables of any type (categorical, continuous, or combinations), which is why it is preferred over a Latent Class Analysis (LCA) in this study (Magidson & Vermunt, 2005).

4.2.1. Indicator Selection

The required data for conducting the LPA consists of selected user-level indicators in the Cryptex exchange dataset, driven by SQ1. Currently, no consensus exists on the number of indicator variables that should be included in an LPA model, although Wurpts and Geiser (2014) mention that incorporating more indicator variables generally leads to better results. Moreover, sufficient indicator variety and both a clear definition of the study population and its size are essential for LPA to guarantee model stability and capture the user heterogeneity (Spurk et al., 2020; Weller et al., 2020). Therefore, it is considered essential that researchers properly guide their choice for specific indicator variables, as a theoretically grounded rationale improves the classification, interpretation and practical usefulness of the resulting profiles (Weller et al., 2020). The indicator selection process follows a two-step logic: first, indicators are conceptually motivated based on existing literature in Chapter 5; second, their empirical

characteristics are assessed through exploratory data analysis in Chapter 6, to ensure feasibility and interpretability within the Cryptex dataset. By doing so, a suitable study population and indicator variables are established, as we deal with an empirical dataset we were not able to construct a dataset beforehand that accompanies LPA best practices.

4.2.2. Model Estimation Procedure

The analysis was conducted in the statistical software *LatentGOLD version 6.0*. A default Maximum-Likelihood estimator was applied, following recommendations by Weller et al. (2020) for researchers new to LPA. For this estimation, a maximum of ten classes was considered as an upper bound to prevent overfitting and model instability and preserve parsimony, supporting the practical interpretability. Given the sample size and the fact that the number of free parameters rapidly increase with higher class models, more than ten classes would likely result in very small classes and unreliable parameter estimates.

Although LPA provides transparent grouping, it is susceptible to its final model selection, and profiles need to be interpreted carefully (Nylund-Gibson et al., 2007; Spurk et al., 2020; Williams & Kibowski, 2016). The optimal number of profiles has to be decided by the researcher based on both theoretical and statistical criteria (Weller et al., 2020). In this study, two criteria are taken into account. First, several relevant statistical information criteria will be compared, with particular inclusion of the Bayesian Information Criterion (BIC) as it is considered the most reliable indicator of model fit (Nylund-Gibson et al., 2007). However, since a relatively small study population was obtained ($N < 300$), attention was also given to the Akaike Information Criterion (AIC), in line with prior recommendations in LPA literature (Morgan, 2015; Nylund-Gibson et al., 2007; Sinha et al., 2021). Class size adequacy was also taken into account to avoid solutions that include very small, mainly instable subgroups. Group sizes below $< 5\%$ were treated extra carefully (Nylund-Gibson & Choi, 2018; Sinha et al., 2021). Entropy is considered informative, but not as a definitive selection criterion as discussed by Sinha et al. (2021), as overfitting models can have large entropies as well.

In addition, the stability of the model was assessed by using different starting values and replications. To assess model stability, different sets of random starting values (100 and 250) and replications (500 and 1000) were analyzed. Even though model stability is not always mentioned in earlier studies that use LPA as a research method (Spurk et al., 2020), it was considered important to review to avoid local maxima and unreliable solutions. When additional profiles yield in small classes and are not consistently replicated across different sets of starting values especially for a small sample size, this raises concerns whether these profiles reflect stable population structure or sample-specific noise (Nylund-Gibson & Choi, 2018). Given the exploratory nature of this study and an initial application of LPA some degree of instability is acknowledged but not treated as disqualifying, but rather an informative caveat for the interpretation of the profiles.

Lastly, two sensitivity analyses were conducted to assess model stability regarding the presence of presumed internal users (further elaborated on in Chapter 6) and severe outliers.

4.3. Post-Hoc Analysis

The latent profiles are constructed on behavioral user indicators which capture the structuring of transactions. However, money laundering risk can be considered a multidimensional concept as explained in Chapter 2, which extends beyond such internal exchange transaction patterns alone. In line with Fu et al. (2025), this limits the ability to fully interpret the identified behavioral profiles, because similar transaction patterns may reflect different underlying strategies or intentions. The authors therefore argue that incorporating contextual information could strengthen analyses and improve interpretability. Accordingly, a post-hoc analysis is performed within this thesis on variables not included in the profile estimation that provide an indication on the positioning within the broader, (digital) financial network and variables associated with higher money laundering risk levels. This may strengthen the assumption of criminal intent associated with specific profiles.

This post-hoc analysis, often referred to as distal outcome analysis in LPA literature (Clark & Muthén, 2009; Nylund-Gibson et al., 2019), serves two objectives. The first is assessing the convergent validity: if the profiles systematically differ on variables that were not used in their construction, this provides supporting evidence for the meaningfulness and interpretability of the obtained profiles (Chin & Yao,

2014), in line with associated benefit of a mixed-method research approach (Jick, 1979). The second objective includes substantive enrichment, which aids interpretation from a law enforcement perspective and provides a more contextual understanding of the users the profiles represent.

To this end, the analysis evaluates whether observed behavioral differences are consistently reflected across multiple data domains reflecting risk potential, thereby strengthening the interpretability of the profiles. Collectively, the approach is guided by the hypothesis that user profiles that not only differ in their transactional behavior, but also exhibit consistent differences across other relevant risk dimensions are considered meaningful segments.

The post-hoc analysis includes three distinctive components, which are further elaborated on in the following subsections. First, internal address-sharing patterns between users are explored to determine if profiles systematically differ in presence and position within this internal network, which can give an indication on the interconnectedness between users and potential clustering around shared services or coordinated behavior.

Second, an illicit exposure screening of transactions using *Chainalysis* is conducted, to identify whether transactions associated with the identified profiles can be linked to known illicit entities or high-risk counterparties.

Lastly, the distribution of internal KYC characteristics across profiles is explored, providing an indication of how users differ in their registered identity and compliance-related attributes within Cryptex.

For each of these three post-hoc components, the results are analyzed by determining to what extent specific characteristics are either overrepresented or underrepresented within a given profile compared to the overall user population. The focus therefore lies on identifying descriptive associations between profile membership and the analyzed characteristics, rather than establishing causal relationships.

For this approach, a classify-analyze procedure is followed, using the modal class assignment to assign each user to a single profile. Although this is not a preferred procedure due to potential misclassification bias, it was considered appropriate given the particularly high entropy of the model (0.96) and the exploratory purpose of this study. More importantly, literature does not consider more extensive procedures such as Bolck-Croon-Hagenaars (BCH) necessary for entropies > 0.8 (Clark & Muthén, 2009).

It is important to note that the analyses in this thesis are exploratory in nature. Since the data originates from a single seized exchange platform and money laundering remains a multidimensional phenomenon even after post-hoc analysis, findings should be treated as exploratory patterns rather than definitive conclusions. Additionally, LPA in general may produce profiles that differ considerably in size (Nylund-Gibson & Choi, 2018). Findings for smaller profiles are therefore interpreted with extra caution throughout this study, also due to the simplified classify-analyze approach.

4.3.1. Network Analysis: Address-Sharing Patterns

In answering SQ3, address-sharing patterns are explored to investigate whether certain profiles show higher levels of address-sharing patterns, as well as in importance and connectivity, by means of a network analysis. This is assumed to provide an indication on on- and off-ramp strategies of users, which may differ based on the obtained profiles.

Network analysis is considered particularly relevant in the field of research focusing on cryptocurrencies, especially for forensic investigations and research seeking to understand behavior within socio-technical systems (Tovanich et al., 2025). By mapping nodes (representing users), edges (representing shared addresses) and applying network metrics, this analysis aims to reveal how specific user profiles may interact within the — in this study — internal address-overlap network.

Here, address-sharing patterns are internally available in the empirical dataset and provide indirect insights into structural and behavioral relationships between users. However, these connections do not imply direct interaction as further outlined in Chapter 8, but can complement the profiles and can be a proxy for shared infrastructural use or behavioral overlap. Such insights are relevant for law enforcement, as they can support the identification of atypical patterns or help prioritize targets for investigation within cryptocurrency exchanges suspected of facilitating illicit activity.

Data Filtering and Network Construction

The analysis is executed in *Python*, using the package `networkx`. First, a network is created based on shared address information. Separate graphs are constructed for deposit and withdrawal address-

sharing, in order to preserve the functional distinction between incoming and outgoing fund flows. This is because deposit addresses reflect the origin of incoming funds while withdrawal addresses reflect the destination of outgoing funds. Although both transaction types were combined for the LPA indicators, the network analysis distinguishes between them as they represent different behavioral mechanisms. Withdrawal-based networks reflect active user choices, potentially related to certain cash-out strategies (Levi & Soudijn, 2020), while a combined network would reflect general address overlap without clear user intent. This distinction is considered important for interpreting the relational patterns between users.

The selection of network metrics requires theoretical justification (Morrison et al., 2022), as the informative value of certain metrics can vary depending on the network type and the specific patterns aimed to investigate (Scott, 2000). Therefore, the theoretical justification will be dependent on the construction of the network itself and will be substantiated in Chapter 8.

Edges are solely based on identical address information shared by users, except for empty, single-character fields, or inputs mentioning "test" (accounting for multiple language specifications). As a result, an edge represents address reuse across accounts and therefore does not have to indicate direct ownership of funds. For example, users may share withdrawal addresses due to the use of the same services, such as cryptocurrency mixers. Edge weights are included and represent the number of unique withdrawal wallets shared by each pair of users.

4.3.2. Chainalysis: Illicit Exposure Assessment

As a second post-hoc component, the first hop addresses of deposits are compared to open-source blockchain data to conduct an initial external risk assessment of transactions associated with each profile. These first-hop addresses represent the first external addresses to which funds are sent or received outside the exchange.

Chainalysis is used to perform a risk-assessment of the first-hop transaction, to explore whether certain profiles exhibit higher levels of affiliation with known illicit activity, which is called its *direct exposure* (Chainalysis, 2023), as explained in section 2.5.

Deliberately, only direct exposure is analyzed, as this provides the highest level of attribution certainty within the scope of this research and is in line with the intended contribution of this specific analysis Table 4.1.

Indirect exposure introduces additional layers of transactional separation as the number of hops considered increases, which leads to a substantial higher risk of false positives as the causal link between the source and the observed (illicit) funds is weakened. Accurately assessing indirect exposure would require more advanced and extended tracing methodologies, which is considered to fall beyond the scope of this thesis.

In line with this argument, the analysis is limited to receiving, deposit-related exposure, focusing exclusively on funds directly received by accounts within Cryptex. This distinction is critical, as for withdrawal transactions it is harder to determine the meaning of direct exposure as explained in Chapter 2. This is because outgoing transactions are inherently linked to Cryptex itself, meaning that those addresses automatically appear as directly exposed to a non-KYC entity. Moreover, such exposure provides limited interpretive value, as it remains unclear whether these recipient addresses are controlled by the user themselves, belong to intermediary services, or represent third-party entities such as exchanges, payment processors, or shared infrastructure providers. Additionally, contrary to deposits, outgoing transactions do not necessarily indicate ownership or direct control over the destination address. This relation is more strongly assumed vice versa. Because Chainalysis does not have complete attribution information on all external blockchain addresses, analyzing withdrawal-related exposure could introduce more ambiguity. In contrast, deposit-related exposure is considered more informative and appropriate for the scope of this thesis, as depositing funds into Cryptex requires a more direct and intentional interaction by the account holder. By restricting the analysis to the receiving exposure of deposit addresses, a higher degree of attribution certainty and interpretability are maintained when characterizing the behavioral profiles.

Exposure data was retrieved on April 21st, 2026. For the analysis, Cryptex vouchers and privacy coin transactions involving Monero were excluded. Specifically, the Bitcoin, Ethereum and Tron blockchain were analyzed. Additionally, no distinction was made between types of illicit activity. Therefore, direct

receiving exposure is operationalized and analyzed as a binary variable for each user.

4.3.3. Conditional Profile Distributions for Attribution Mitigation Behaviors

The final component of the post-hoc analysis examines the profiles their relationship to proxies of both identity attribution mitigation and financial attribution mitigation strategies (Miedema et al., 2023). These represent two fundamentally different layers at which attribution can be mitigated, which are often associated with illicit intent.

For this research, KYC-compliance and the use of privacy coins are used as proxies to represent respectively the identity layer and the transaction layer. Active KYC-avoidance and the use of privacy-enhancing cryptocurrencies are commonly discussed in literature as important mechanisms that contribute to these strategies (Financial Action Task Force, 2020).

Moreover, both factors are directly observable in the dataset and match common investigative filtering strategies. This makes them suitable for a structured post-hoc analysis. Therefore, they offer a meaningful way to explore the relation between attribution mitigation strategies and the constructed profiles, without trying to cover all possible obfuscation techniques.

KYC-compliance

Finally, as a component of the post-hoc analysis, KYC-compliance is analyzed based on the assumption that certain profiles may benefit, given their intent and associated transactional behavior, from either completing or avoiding the KYC-process. It should be noted, however, that the exact implementation of KYC-procedures, identity control and subsequent approval by Cryptex remains unknown, which may affect the validity of conclusions drawn from this analysis. Yet, it is considered that the intent to apply, since it was not a mandatory process for registering and operating an account (United States Attorney's Office, 2024), still provides valuable information.

Use of Privacy Coins

As explained in Chapter 2, privacy coins provide enhanced secrecy on the pseudo-anonymous nature of virtual currencies, and as a result are often preferred by criminals (Fu et al., 2025). For this analysis, a straightforward descriptive analysis is conducted to see which profiles show higher affiliations with users who have conducted at least one transaction involving privacy coin Monero on the platform. This allows for an initial exploration of whether the use of privacy-enhancing assets is associated with specifically with the constructed behavioral profiles.

4.4. Interviews with Law Enforcement

Finally, in relation to SQ4 and SQ5, a focus group was conducted together with a homogeneous group of experienced investigators and analysts to validate and contextualize the findings from a law enforcement perspective. This part addresses the earlier critique on existing research and the limited involvement of investigative practice. The focus group functions as a supportive rather than a primary component of this research within this mixed-method study, as outlined in section 4.1.

The focus group serves three objectives. First, it provides expert validation on the extent to which the obtained profiles are perceived as realistic. Second, it assesses the value and practical applicability of LPA for their operational practices. Third, the focus group aims to obtain a critical reflection on potential gaps or alternative interpretations that are important for deriving the conclusions of this research.

Advantages of this method include that multiple experts can be interviewed simultaneously, making it well-suited to the scope of the qualitative component within this thesis, as well as its exploratory objective. Additionally, the group dynamics between respondents during the focus group are assumed to automatically help surface the most important issues, while also providing a form of quality control, as respondents can refine, challenge, or build upon each other's statements (Leung & Savithiri, 2009).

4.4.1. Focus Group Structure and Analysis

The focus group is semi-structured in nature, which allows to ask follow-up questions to further dive in to aspects deemed relevant or unexpected during the discussion (Galletta & Cross, 2013). For data collection, the session is audio recorded with a secured device, after which a focus group summary is written, and thematically analyzed to conclude on the results. Ethical considerations, next to those

of section 4.6, were also upheld during this part of the research by agreeing on an Informed Consent, which among other things includes that participation is voluntary, responses are kept anonymous and findings are reported at the group level.

Beforehand, participants are provided a hand-out with a summary of the profiles, after which a moderated discussion takes place following a semi-structured topic list derived from the research findings. An overview of the proposed questions, as well as the further protocol can be found in Chapter 9. Respondents were deliberately not informed about the post-hoc profile analysis, as both approaches are designed to complement each other in the validation of the profiles. Providing this information beforehand could potentially bias experts by influencing the interpretation of the profiles.

Respondents

For the focus group, four experts were interviewed. No specific sampling method was used, as respondents were selected pragmatically based on their expertise and availability, rather than through a formal sampling strategy. The respondents had varying levels of experience and knowledge related to cryptocurrency investigations and data analysis. The group included experts from both general investigative teams (AOT), who are primarily involved in initial data assessment and operational screening, and more specialized investigators with advanced analytical expertise. All respondents were employed within the same law enforcement organization, with work experience ranging from 2 to 6 years, all of them work weekly to every day with transactional data. An overview of respondents can be found in Appendix E.

The size and selection of the respondent group was considered adequate for the scope of this qualitative component within this thesis, as its objective is to mainly get an initial assessment on the obtained results and discover the range of perspectives. The inclusion of both general and more specialized respondents, allowed for a broader perspective on the obtained profiles and the practical value they might bring, depending on the kind of investigative research conducted.

4.5. Research Flow Diagram

Figure 4.1 on the next page summarizes this chapter by illustrating the flow of this research, linking sub-questions, methods, and chapters.

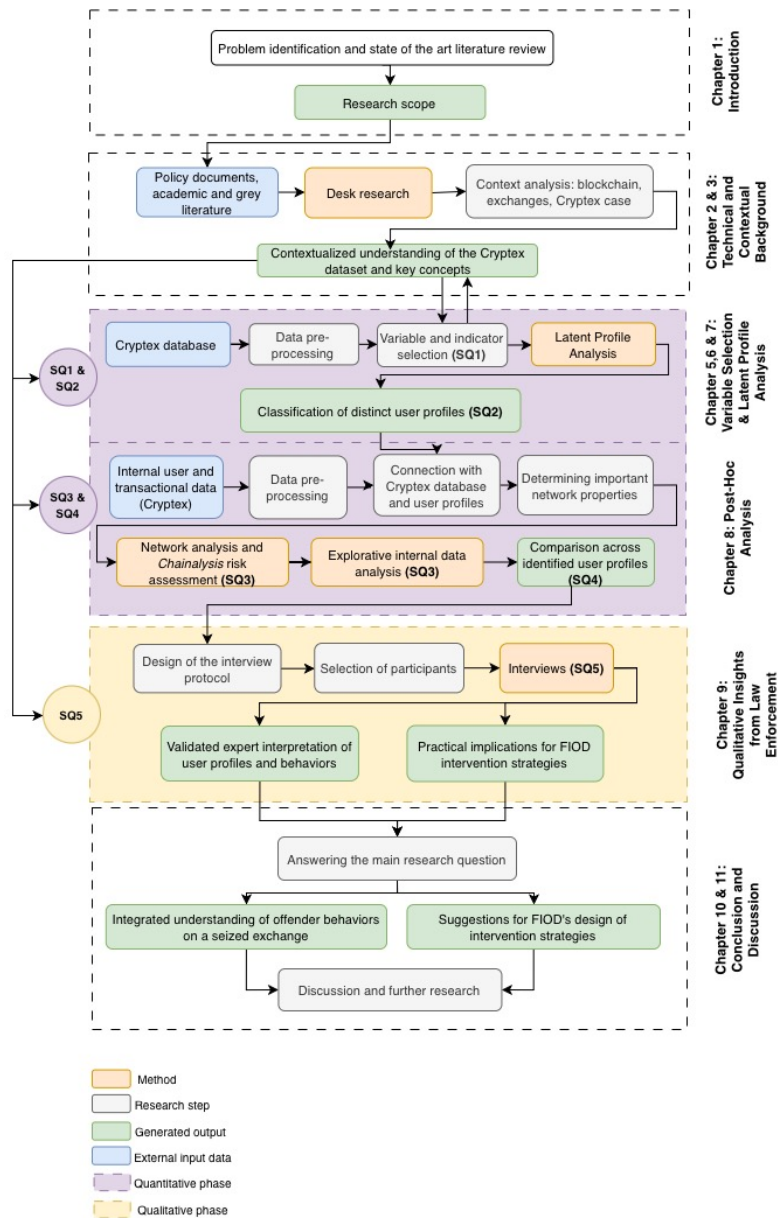


Figure 4.1: Research Flow Diagram

4.6. Ethical Considerations

Data used in this research has been provided in correspondence with the Dutch FIOD and is analyzed anonymously, to minimize the risk of user attribution. Additionally, this thesis has been reviewed by the FIOD to ensure it does not interfere with any ongoing investigations. For the conducted focus group, a Human Resource Ethics Checklist (HREC) was completed, together with the development and implementation of a data management plan and informed consent protocol. In line with the law enforcement context in which this research was conducted, all interpretations and methods are considered to remain aligned with investigative purposes and proportionality. Findings are presented at the profile level, rather than the individual level, to ensure the analysis cannot be used to trace or identify specific individuals. Moreover, the investigative recommendations that can be derived must sufficiently account for the limitations of this study.

5

Conceptual Lens for Behavioral Indicators

This chapter aims to answer SQ1: “*What types of user behavior can be identified as potentially meaningful for distinguishing behavior within crypto-exchange environments?*”. The outcome of this chapter is a conceptual indicator framework that provides theoretically motivated behavioral dimensions which serves two purposes: it provides conceptual legitimization for the subsequent Latent Profile Analysis (LPA), and it offers a structured foundation for future latent class modeling in similar contexts.

5.1. Behavioral Domains of Cryptocurrency Exchange Usage

For the construction of this framework, the focus is put on behavior performed by users *on* cryptocurrency exchange platforms. These actions may provide an indication of differences between legitimate and illegitimate users, as certain behaviors may contribute to a higher level of anonymity by for example, affecting the traceability of funds. Following Miedema et al. (2023), a distinction can be made between *financial attribution mitigation* (obscuring the traceability of funds) and *identity attribution mitigation* (obscuring the link between a person and an account).

This research focuses on observable financial activity *within* the exchange environment that structure users’ transactions over time and may be consistent with financial attribution mitigation. Therefore, the analytical focus is put on operational transactional practices rather than technical security operations to obscure identity, such as routing strategies or the decision to select a specific exchange (for example, choosing a DEX over a CEX, or selecting an exchange based on its jurisdictional location). Moreover, this is conceptually distinct from other security practices or contextual indicators, such as exposure measures and KYC-related characteristics. While these variables may provide important information on risk and investigative relevance, they do not directly indicate how users structure their transactional activity on the platform, they provide contextual information regarding the environment, attribution, or potential risk associated with that behavior. Therefore, they are considered separately from the behavioral dimensions developed in this chapter.

This distinction is followed deliberately, as they also represent different behavioral domains. Combining them would result in a heterogeneous construct, which is methodologically undesirable as LPA assumes that indicators are manifestations of a single underlying latent structure (Weller et al., 2020). Previous research including the use of LPA within comparable contexts, such as van de Laarschot and van Wegberg (2021) and Miedema et al. (2023), mainly examined security-related practices related to identity attribution mitigation. This thesis thereby contributes theoretically to analyze whether latent user profiles also emerge based specifically on financial behavior. Furthermore, financial behavior provides more direct and observable financial traces that are measurable within exchange data, making it particularly relevant for law enforcement and the context of this research.

However, it is important to note that financial and identity-related security practices may interact in practice (Financial Action Task Force, 2020; J5, 2024), also represented in Figure 5.1. For example, routing

strategies may be properly arranged before conducting any transactions, or the decision for a specific exchange may depend on the desired transaction size and KYC-requirements. Financial behavior on the other hand, may at times serve security-related purposes, such as intentionally fragmenting transactions into smaller amounts may be a deliberate strategy to reduce detectability (Financial Action Task Force, 2020). However, the LPA remains neutral regarding the user's intent and thus this study focuses solely on observable and measurable (transactional) patterns.

Accordingly, a conceptually homogeneous indicator set is prioritized to ensure interpretability and methodological appropriateness. As a result, latent behavioral profiles are conceptualized as emergent patterns of operational transactional practices, while acknowledging that such patterns may be influenced by broader strategic contexts.

5.2. Towards a Conceptual Framework

As discussed in the previous section 5.1, this study focuses on the domain of operational transactional practices, which can be further divided into behavioral dimensions that capture structural aspects of transactional behavior on exchange platforms.

These dimensions are informed by policy literature from J5 (2024) and Financial Action Task Force (2020) that categorize transactional features that influence traceability and attribution. These are used to identify relevant dimensions of transactional behavior within cryptocurrency exchanges. Together, these dimensions capture the main structural properties of exchange usage and the structuring of transactions on such platforms: how much users transact, in which direction value flows, how monetary value is organized, how activity is distributed over time, and how broadly activity is spread across assets.

5.2.1. Operational Transactional Practices: Behavioral Dimensions

Activity intensity

First, activity intensity can be distinguished. This dimension captures the frequency of interaction with the exchange platform and therefore reflects the extent to which the exchange is actively utilized.

Flow Directionality

This concerns the relative balance between in- and outflows within the exchange. As a result, this dimension captures whether the exchange usage is oriented towards asset accumulation or asset outflow. In contrast to the activity dimension, this reflects a more directional orientation of financial flows.

Monetary Structure

Third, monetary structure reflects how monetary value is organized within the user's activity. It includes both magnitude and variability, for example if large or small transactions are performed and whether these are variable or stable amounts. Within this dimension, monetary entry positioning is also considered following Financial Action Task Force (2020), as the magnitude of the initial deposit may reflect the scale of initial financial engagement within the exchange environment

Temporal Structuring

This dimension captures how activity is distributed over time. For example, it represents whether behavior is performed in concentrated periods of activity, or if it is more sustained.

Asset Diversification

Asset diversification indicates the extent to which exchange activity is distributed across multiple cryptocurrencies, thereby it captures if users concentrate their activity in a limited number of assets or engage with a broader range of tokens. Asset diversification may reflect the breadth of engagement by differences in trading strategies, operational complexity, or attempts to distribute value across multiple assets.

Table 5.1 summarizes these five behavioral dimensions that capture the core structural properties of financial exchange behavior and provide the foundation for the subsequent operationalization.

Table 5.1: Operational behavioral dimensions within exchange usage

Dimension	Structural property	Core Question
Activity intensity	Frequency	<i>How much exchange activity is performed?</i>
Flow directionality	Direction	<i>What is the relative balance between inflows and outflows?</i>
Monetary structure	Magnitude and spread	<i>How are monetary values structured within exchange activity?</i>
Temporal structuring	Timing	<i>How is activity distributed over time?</i>
Asset diversification	Breadth	<i>Across how many assets is activity distributed?</i>

5.2.2. Conceptual Framework

Figure 5.1 depicts the conceptual framework used for this thesis. Exchange-based behavior is conceptualized as consisting of multiple behavioral domains. As discussed, while operational security practices and operational transactional practices may interact in practice, the LPA is exclusively based on indicators that capture patterns in how users structure transactions on the exchange within this research. As shown, operational transactional practices are represented through five structural behavioral dimensions, which are subsequently operationalized into measurable indicators in the following chapters.

It should be noted that this conceptual framework does not provide an exhaustive overview of all the operational practices and contextual indicators. Instead, it provides a conceptual foundation that guides the argumentation and analytical focus of this research.

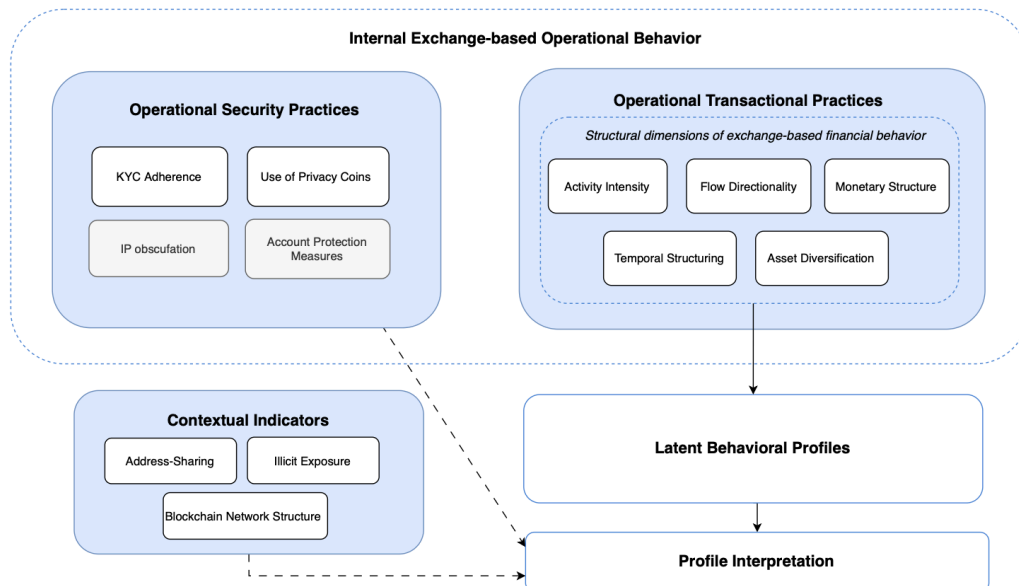


Figure 5.1: Conceptual framework of behavioral domains and the identification and their relation to the latent behavioral profiles

5.3. LPA: Data and Selection Criteria

Based on the above, the final indicator selection for conducting the subsequent LPA is guided by a set of principles to ensure that the latent user profiles capture meaningful and interpretable patterns of behavior. These principles are informed by two sources: (i) the user-centered and exchange-specific operational level outlined in section 5.2, and (ii) methodological guidance from the latent class and latent profile analysis literature.

1. **Availability within the dataset and population:** It is important that the chosen indicator variables, representing the operational behavioral dimensions, have been recorded within the dataset for the complete, selected study population. Unless a missing value represents absence of behavior, which can be included. However, it is important to note the sensitivity of LPA to the presence

of many zero values, so-called 'zero-inflation' (Van Lissa et al., 2024).

2. **Data quality:** Critically, indicators must be reliably observable within the dataset. This requirement is inherently dependent on the nature and structure of the available data. Variables with substantial missingness or unreliable recordings are excluded to ensure robustness of the analysis. Therefore, data preparation should precede the specification of the final conceptual model.
3. **Measurability on the user level:** LPA is a person-centered methodology instead of a variable-centered analysis (Muthén & Muthén, 2000; Weller et al., 2020). This is relevant for reliable estimation and subsequent classify-analyze approaches. In law enforcement contexts, the goal is usually not only distinguishing between groups, but also retain the ability to examine individuals within those groups for potential attribution.
4. **Coverage:** Indicators are selected in such a way that they collectively cover a broad range of the behavioral patterns of interest, here discussed in section 5.1. This ensures that multiple dimensions of user behavior are represented and they are therefore related (Spurk et al., 2020).
5. **Sufficient variability:** Chosen indicators must exhibit adequate between-user variability. Variables with limited variance do not provide enough capacity for an LPA to meaningfully capture differences between latent profiles.
6. **Non-redundancy:** Although coverage is important, indicators should not measure the same underlying construct and therefore be distinct (Spurk et al., 2020). Conceptually overlapping or correlated variables should be avoided or removed to prevent redundancy and improve interpretability.
7. **Interpretability:** Partially influenced by the last principle and guiding their operationalization, interpretability of the indicators is prioritized. Instead of reflecting purely statistical groupings, in this way meaningful conclusions on user behavior can be derived from the profiles instead.
8. **Generalizability:** As this research also aims to contribute methodologically, indicators should preferably reflect behaviors that are observable across different cryptocurrency exchanges. This enhances the feasibility and replicability of this research, as well as generalizability of the findings across multiple exchanges.
9. **Number of indicators:** Currently, there is no established consensus on the optimal number of indicators in LPA. Although a larger set of variables may allow for more detailed profile differentiation, it simultaneously increases model complexity and the risk of overfitting (Wurpts & Geiser, 2014). Moreover, research indicates that the adequacy of the model does not only depend on the sample size, but also the number of indicators, where more complex models required substantially larger samples for stable and reliable solutions (Nylund-Gibson et al., 2007). Consequently, the number of indicators should be aligned with the available sample size and overall model complexity.

These criteria form the foundation for the selected indicator variables for the LPA in section 7.1. However, criteria 1, 2, 3, 4 and 5 are bounded by the available empirical dataset. Therefore, first in Chapter 6 an exploratory data analysis is performed to define a relevant study population and variables that suffice the criteria of this section.

6

Empirical Assessment of Behavioral Indicators

While the previous chapter provided a conceptual framework on the behavioral dimension considered for this research, this chapter offers empirical support by examining whether these behavioral indicators are sufficiently prevalent in the dataset and whether they meet the selection criteria defined earlier. Based on this, this chapter defines the study population for this research by outlining and inspecting the dataset in further detail.

These findings inform the selection, interpretation and preparation of indicators and definition of the study population which functions as a necessary step and foundation for the Latent Profile Analysis (LPA) in the subsequent Chapter 7.

6.1. Dataset Overview and Scope

Based on the conceptual framework of Chapter 5, it can be concluded that transactional user activity is a key dimension for distinguishing behavioral patterns, including potentially criminal behavior, therefore serving as the central focus of this analysis. Consequently, particular emphasis is put on data that represent transactional activity within the empirical dataset. As this research aims to construct a LPA to identify distinct user groups based on their behavior within the exchange, the analysis focuses specifically on *user-level* and importantly, *user-initiated* activity. This scope reflects the objective of capturing behavioral heterogeneity and distinguishing user types on the platform that can aid law enforcement operations, rather than finding (other) statistical patterns that relate to processes on a different abstraction-level. User activity is captured through a unique identifiable ID assigned to each user in the dataset.

Based on these considerations, the dataset is inspected for *user IDs with ≥ 1 financial transaction* found in any of the transaction-related tables within the dataset. The population under study is hereafter referred to as the *financially active population*, which is seen as an important step in preparation for conducting an LPA (Sinha et al., 2021). Externally, supplementary data, such as blockchain information, is discussed and considered conceptually relevant in the previous chapter, but empirically out of scope for this LPA and is part of the post-hoc analysis in Chapter 8. This restriction is adopted to enhance both the practical applicability for law enforcement, where access to external data may be limited, as well as replicability from an academic viewpoint.

As mentioned in Chapter 3, Cryptex underwent a website migration, in which it also changed its transactional structure. As a result, to ensure behavioral comparability, methodological consistency and completeness of the data, the analysis is restricted to data collection after this migration. This implies a focus on data collected between March 2023 and its formal takedown in September 2024.

6.2. Population Structure and Coverage

This section discusses the user base and relevant subpopulations in the dataset, for motivating the selection of the study population usable for the LPA.

6.2.1. General User Population

Although the platform registered more than 20,000 unique user IDs, in contrary to research expectations, only a limited subset actively engaged in transactional activity recorded in the dataset as depicted in Table 6.1 and this table is further decomposed in the following subsections.

Table 6.1: Overview of the user population

	Number of unique <i>user_ids</i>
Total population	20,416
Financially active population	261
<i>Deposit and/or withdrawal users</i>	258
Internal users	52

Based on Table 6.1, it becomes clear that the majority of users in the dataset are not part of the financially active population. Two main reasons account for this. First, to preserve data purity and to ensure validity of results, only data covering the period from March 2023 to September 2024 is included; older records starting July in 2017 that predate the website migration in 2023 have been excluded, despite containing potentially a large number of users. Second, it is likely that a large share of users created an account but never actually engaged in any transactions. This is a pattern not considered uncommon, given that VASPs have been known to pose financial risks to their users (Conlon et al., 2023), thus trust is an important factor in making use of such services (Toufaily & Zalan, 2024).

Furthermore, according to internal FIOD investigative information, a substantial share of users without recorded transactions may have been active through other platform services that were not included in the available data. As a result, these users cannot be considered fully inactive, nor can their underlying behaviors be reliably deduced. Consequently, this group is excluded from the behavioral analysis, as assigning generalized labels (e.g., “window shoppers”) would be speculative and analytically unsupported.

6.2.2. Financially Active Population

As mentioned, we consider the financially active population as the user IDs with ≥ 1 financial transaction(s). Tables capturing activity associated with this population are *core_order*, *core_transaction*, *core_cryptotransaction*, *core_simpleexchange*, *core_withdrawal*, *core_deposit*, *core_exchangerpayment*. Given the focus on user behavior, datasets that reflect balances or other system states, rather than user-initiated events or actions, are excluded. Table 6.2 depicts the descriptive statistics respectively for each table. It is important to note that *core_exchangerpayment* does not contain unique IDs, but e-mail addresses. These were linked to existing *user_ids* by identifying the overlap with the e-mail addresses in the *auth_user* database, containing all IDs. This yielded in 220 matched user IDs. However, it cannot be verified with certainty whether identical e-mail addresses correspond to the same underlying user, which limits the usefulness of this specific dataset, as reliable attribution of behavior to a specific user is considered essential.

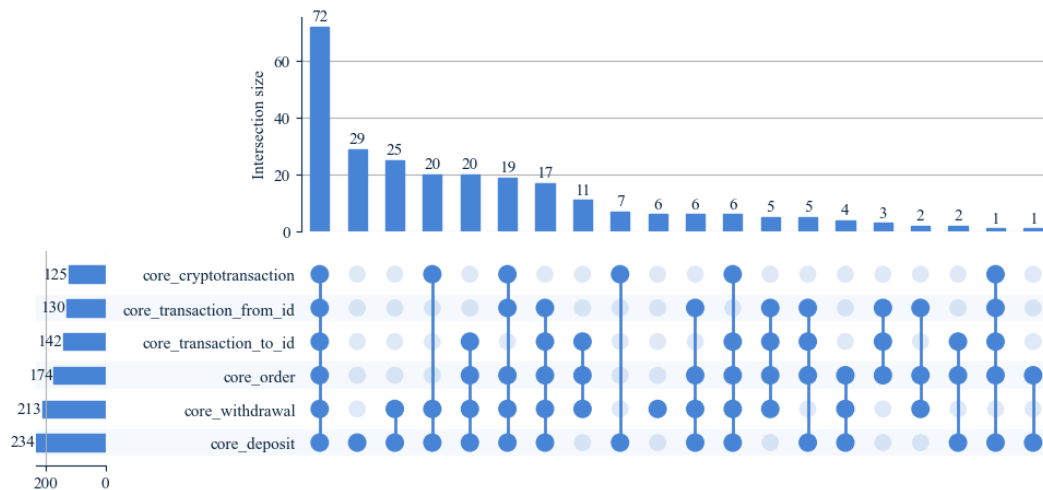
Table 6.2: Descriptive statistics per transactional table

Table	N Unique users	% of total <i>user_ids</i>	N Transactions	Mean per user	Median per user	SD per user
<i>core_order</i>	174	0.85%	5,204,285	29,909.7	11	287,845.6
<i>core_transaction</i>	169	0.83%	6,132	19,123.5	22	161,668.7
<i>core_cryptotransaction</i>	125	0.61%	336,282	2,668.9	4	24,207.4
<i>core_simpleexchange</i>	100	0.49%	4,987	49.9	4	277.0
<i>core_withdrawal</i>	213	1.04%	37,304	175.1	4	850.1
<i>core_deposit</i>	234	1.15%	170,412	728.3	5	9,403.0
<i>core_exchangerpayment</i>	*220	1.08%	6,132	—	—	—

***Note:** Unique users in *core_exchangerpayment* should be interpreted indicatively, as it does not contain unique user IDs, but e-mail addresses, as it can not be verified that identical e-mail addresses belong to the same underlying user, which imposes additional uncertainty.

Based on the mean and median values per user, it becomes clear that all datasets, the number of transactions per user is highly right-skewed, indicating significant outlier dominance.

Moreover, Table 6.2 does not indicate the prevalence of unique user IDs across the dataset. Therefore, Figure 6.1 illustrates the coverage of unique user IDs over the different financial tables.

**Figure 6.1:** User coverage across transactional tables

As depicted by 6.1 the spread of user IDs is highly uneven across tables. The *core_simpleexchange* is not depicted in Figure 6.1, as all users recorded in this dataset fall entirely within the *core_order* population and therefore do not add additional user coverage. Also *core_transaction* is split in a *to_id* and *from_id*, reflecting the structure of the dataset itself. Finally, although partially *core_exchangerpayment* could be linked to existing user ID e-mail addresses, this imposes uncertainty at the user-identification level, which is therefore not presented in Figure 6.1.

Most of the 261 users considered as 'financially active' can be recorded in 258 unique user ids from withdrawal and deposit, making these two the largest representable tables.

6.2.3. LPA Sample Construction

First, indicators derived from the conceptual framework of Chapter 5 must be reliably available in the dataset and measurable across the study population. Based on the user analysis in this section, it can be concluded that the dataset captures multiple types of transactional operations, yet with a strongly uneven distribution of users across these operation types. As a result, only a limited subset of users can be considered financially active for certain transaction categories.

This uneven distribution poses a challenge for conducting an LPA, as it increases the risk that the model primarily reflects *observed* differences in transaction category activity (e.g., *core_order*, *core_transaction*,

core_cryptotransaction), rather than capturing *unobserved* behavioral heterogeneity, which is the core objective of LPA. This concern is further amplified as an LPA is sensitive to zero values in indicators (Sinha et al., 2021). Therefore, inclusion of other dataset tables would introduce a substantial number of inactive accounts, which can result in a zero-inflated population and profiles that are thus trivially distinguished by activity versus inactivity and thus *observable* heterogeneity, rather than by meaningful differences in behavioral intensity or patterns. Moreover, across other transactional tables, user identifiers and relevant indicators cannot be constructed in a sufficiently consistent manner to support a valid LPA, which would also result in table specific indicators and zeros.

In addition, given the limited transparency of the underlying data-generating process, it is essential to prioritize tables that most plausibly reflect user-initiated behavior, rather than system-level or book-keeping processes. Taken together, these considerations motivate the selection of the *core_deposit* and *core_withdrawal* tables for the construction of LPA indicators. These datasets represent a relatively large, homogeneous and comparable user population, which is more suitable for LPA. To provide additional insight into the data, boxplots are presented in Figure 6.2 to visualize the distribution and spread of transactions per user across the full study population.

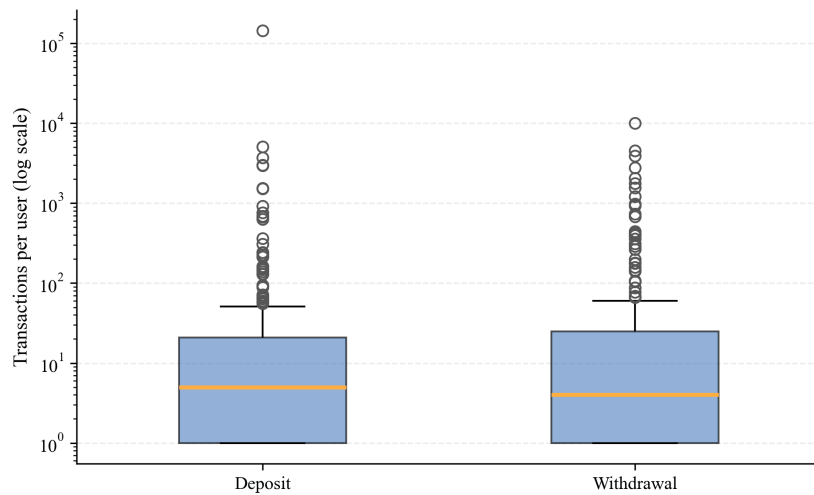


Figure 6.2: Number of deposit and withdrawal transactions per user on a log scale

6.2.4. Presence of Internal Administrative Accounts

Several accounts were identified as presumed internal users, based on the e-mail addresses used. However, it is plausible that more internal accounts were used, for instance for internal transfers or as warm wallets. This makes it hard to establish whether tables really represent individual user intent, or are part of a secondary transactional processing layer of the platform.

In total, 52 internal users across the complete dataset were observed and 15 internal users within the study population, which accounts for 6.2% within these two dataset tables of deposits and withdrawals. Table 6.3 and Figure 6.3 provide an overview of the internal users and their transaction frequency, showing that extreme variability among internal users also exist. Since the median is considerably higher than the total sample population in Table 6.2, generally, it can be concluded that internal users do show higher levels of activity and are therefore carefully considered within the analyses of this thesis.

Table 6.3: Descriptive statistics of transactions per user for deposits and withdrawals (internal users)

Transaction type	Mean per user	Median per user	SD per user
Deposits	903.47	17	1,658.59
Withdrawals	1,603.58	29	3,031.56

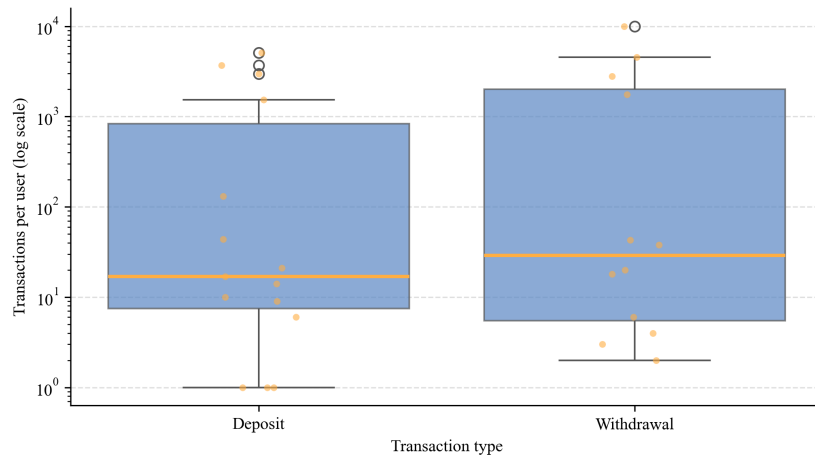


Figure 6.3: Transactions per user (internal users) on a log scale

These presumable internal users are retained in the LPA analysis in order to preserve a complete representation of the dataset, while information on their presence is taken into account by conducting a sensitivity analysis in Chapter 7 and in the interpretation of the resulting profiles and post-hoc findings of Chapter 8.

6.3. Activity Concentration

This section provides further analysis on the sample population stemming from *core_deposit* and *core_withdrawal*.

Additionally, Figure 6.4 shows the distribution of account creation over time across the study population. A large share of accounts (36.8%) was created after the website migration in March 2023, which is expected given that the study population is derived from transaction data available only from this date onwards. This is not a surprising outcome, as the use of digital service infrastructures is largely built on trust and reputation (Toufaily & Zalan, 2024), which needs to be acquired over time. This temporal pattern is relevant when interpreting activity patterns, as it indicates that the sample is not uniformly distributed across the platform's lifespan.

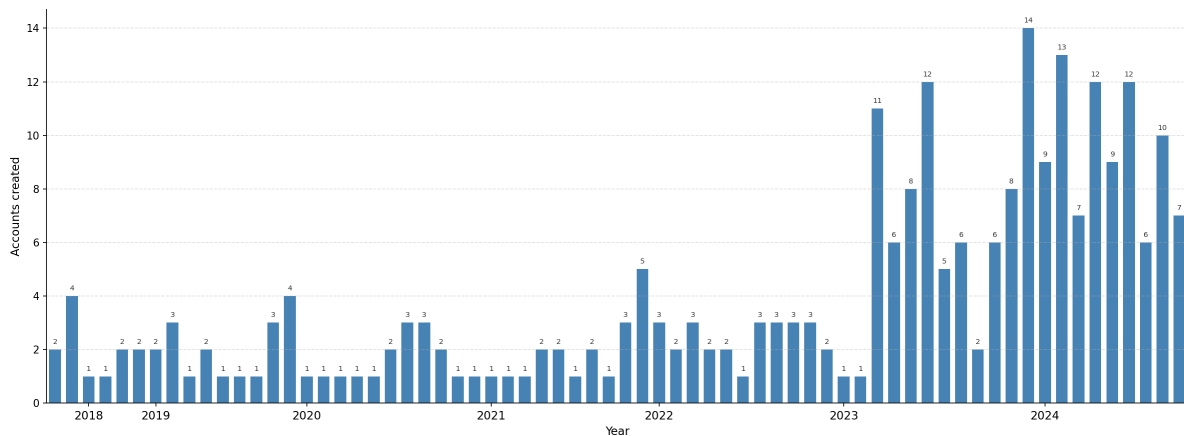


Figure 6.4: Account creation frequency for the study population

6.4. Conclusion and Implications for LPA

Taken together, this supports the decision to use only *core_deposit* and *core_withdrawal* as input for the subsequent LPA. These tables represent a large and homogeneous group of users and account for

the vast majority of the financially active population, which further strengthens their suitability for robust latent profile estimation. However, reflecting on Chapter 3, this constitutes an important methodological decision in which the LPA is only applied to one of the platform's functionalities. Namely, the possibility to open a custodial wallet, as discussed in section 3.2. Given the characteristics of the user population also referring to the website migration, this functionality is presumably not the platform its main selling point.

By redefining this research scope, we obtain a study population that is as clean as possible and minimally affected by any exchange-internal or system-driven events. In this sense, data purity and validity is prioritized over sample breadth. The sample size of 258 constitutes a relatively small sample, which in LPA can lead to convergence issues and unstable solutions. However, according to Nylund-Gibson and Choi (2018) a sample size of 258 users can be considered sufficient, depending on how the model is constructed. For $N < 300$, Henson et al. (2007) found that results are less reliable and a Monte-Carlo simulation to determine the statistical power and reliability of the solution. However, Nylund-Gibson and Choi (2018) and Spurk et al. (2020) mention that these power calculations are very complex and require substantial knowledge on population parameters and specifying the "true" model, which is not always possible due to the exploratory nature of LPAs. As the same holds for this research, the analysis proceeds without a formal simulation. Finally, regarding the exploratory nature of this study, the analysis is bound by the available data and methodological requirements. Accordingly, the estimated latent profiles should be interpreted as behavioral segments within the specific study population, rather than as generalizable population-level typologies. Moreover, indicator parsimony and thus a limited number of indicators is considered necessary given the sample size constraints that have been identified in this chapter (Nylund-Gibson & Choi, 2018; Weller et al., 2020).

7

Latent Profile Analysis on User Types

In the previous Chapter 5, a general theoretical framework was constructed to support the indicator selection used for this method and the data has been empirically reviewed from this conceptual lens in Chapter 6. This chapter presents the Latent Profile Analysis (LPA) conducted to identify distinct user profiles within the available dataset of Cryptex, with the aim of answering SQ2: “*What latent user profiles can be identified within the Cryptex exchange?*”.

7.1. Indicator Selection and Data Preparation

This section discusses the final set of chosen indicators, based on the previous Chapter 6 and section 5.3, as well as the necessary data preparation prior to the LPA. Indicator selection is grounded in the conceptual framework of section 5.2, the empirical assessment of the data (Chapter 6) and criteria formulated in section 5.3.

Accordingly, the final set has been chosen based on their theoretical value (Chapter 5) and empirical suitability (Chapter 6). Particular attention is provided to the criteria coverage, spread, interpretability and redundancy. Additionally, parsimony of the indicators is prioritized as mentioned in section 6.4 due to the sample size (Nylund-Gibson & Choi, 2018), while remaining the core dimensions of transactional operational practices.

7.1.1. Operationalization

The behavioral dimensions outlined in section 5.2 must be operationalized in order to conduct the LPA. Therefore they must be translated into dataset-specific indicator variables. Each selected indicator satisfies the predefined criteria regarding variability, interpretability of section 5.3, and relevance to operational financial practices. Where multiple potential indicators were available for a given dimension, preference was given to variables that demonstrated sufficient variability across users, minimal conceptual overlap with other indicators, but also ensuring model parsimony and interpretability.

All indicators were computed at the user level, where i represents a unique *user_id* in the study population and t the individual transactions. All equations were implemented in *PostgreSQL* database coding, to retrieve the operationalized variables from the dataset.

Number of Deposits and Withdrawals

This equation calculates the total count of deposits and withdrawals per user i , thus resulting in two separate indicators. k represents the transaction type, implying either deposit or withdrawal.

$$N_i^{dep} = \sum_{t=1}^{T_i} \mathbf{1}(\text{transaction}_t = k) \quad (7.1)$$

Table 7.1: Operationalization of Conceptual Behavioral Dimensions

Behavioral Dimension	Operationalization	Conceptual Interpretation
Activity Intensity	Number of deposits (7.1) Number of withdrawals (7.1) Number of active days (7.2)	Captures the overall level of transactional engagement within the exchange.
Flow Directionality	Deposit-to-withdrawal ratio (7.3)	Indicates whether exchange usage is oriented toward accumulation or outward distribution.
Temporal Structuring	Burstiness (7.5) Account lifespan (7.4)	Approximates how exchange activity is distributed across time.
Monetary Structure	Average transaction value (7.6) Transaction amount variability (7.7) Initial deposit amount (7.8)	Reflects the scale and dispersion of monetary flows.
Asset Diversity	Dominant asset share (7.9) Number of distinct assets (7.10)	Represents the breadth of engagement across cryptocurrencies.

Active Days

This equation calculates the number of distinct calendar days on which a user i conducted at least one deposit or withdrawal transaction. $\mathcal{T}_i$ represents the set of transactions of user i , different from the number of transactions T_i .

$$D_i = |\{\text{date}(t) : t \in \mathcal{T}_i\}| \quad (7.2)$$

Deposit-Withdrawal Ratio

Flow directionality is operationalized as the ratio between the number of deposits and withdrawals per user. To avoid division by zero and to reduce extreme values for users with very low withdrawal counts, a smoothing constant of 1 is added to both components:

$$DWR_i = \frac{N_i^{dep} + 1}{N_i^{with} + 1} \quad (7.3)$$

In this equation, N_i^{dep} represents the total number of deposits by user i , and N_i^{with} the total number of withdrawals. Here, values greater than 1 indicate deposit-heavy activity, values below 1 indicate withdrawal-heavy activity, and values close to 1 reflect relatively balanced inflow and outflow behavior.

Account Lifespan

Difference between the first and last observed transaction date for user i .

$$L_i = \begin{cases} \max(\text{date}_i) - \min(\text{date}_i), & \text{if } T_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad (7.4)$$

Since only users i are included that are present in either the withdrawal and deposit dataset or both, no values equal to 0 follow from this operationalization.

Burstiness Coefficient

To capture how activity is spread over time, a slightly transformed burstiness coefficient is computed for each user, as defined by Goh and Barabási (2008). This coefficient represents the spread of user activity over time, adding structure behind the transaction frequencies.

$$B_i = \begin{cases} \frac{\sigma_i - \mu_i}{\sigma_i + \mu_i}, & \text{if } \sigma_i + \mu_i > 0 \text{ and } n_i \geq 3 \\ 0, & \text{otherwise} \end{cases} \quad (7.5)$$

To calculate the inter-event times, the number of days between transaction (deposit or withdrawal) is computed. Thus, for a user with n_i transactions, this results in $n_i - 1$ inter-event times. μ_i represents the mean inter-event time and lastly σ_i the standard deviation of inter-event times. n_i is the number of transactions of user i . The sample standard deviation is used, because it is assumed that the observed transactions are only a part of the user their underlying behavioral process. Moreover, users with less than three transactions do not have sufficient observations to compute meaningful parameters, thus for these cases the burstiness coefficient is set to 0.

The burstiness coefficient lies between -1 and 1 , where a value closer to -1 represents highly regular, routine behavior, while values around 0 approximate random behavior and 1 indicates very clustered behavior, characterized by many short intervals of high activity, followed by longer periods of inactivity.

Average Transaction Volume (USD)

The average transaction value in US Dollars (USD) was computed for the deposit and withdrawals together, to get an impression of the monetary value, making virtual currency amounts comparable. For each user i the general form is:

$$\bar{X}_i^{(k)} = \frac{1}{N_i^{(k)}} \sum_{t=1}^{T_i} |x_{it} \cdot r_{\text{day}(t), \text{cur}(t)}| \mathbf{1}(\text{transaction}_{it} = k) \quad (7.6)$$

Within this equation, k indicates the transaction type: deposit, withdrawal or overall. Conversion to US Dollars was based on daily exchange rates extracted from *CoinMarketCap*, for a cut-off value of 200, for 13 currency types¹. Exchange rates were only determined for currencies that exceeded a cut-off value of 200 in aggregate transaction activity. For all remaining currencies, exchange rates were set to 0 during the conversion procedure. The resulting measure therefore represents the average transaction value in USD. Given the nature of the platform, presence of extreme price speculation is not expected to substantially affect the validity of this conversion. Based on k , this thus results in three indicators of which their intercorrelations were inspected.

Transaction Size Variability

To capture within-user variability in transaction size, the standard deviation of transaction amounts was computed separately for deposits and withdrawals. The indicator is defined as:

$$\sigma_i^{(k)} = \begin{cases} \sqrt{\frac{1}{N_i^{(k)} - 1} \sum_{t=1}^{N_i^{(k)}} \left(x_{it}^{(k)} - \bar{x}_i^{(k)}\right)^2}, & \text{if } N_i^{(k)} > 1 \\ 0, & \text{if } N_i^{(k)} \leq 1 \end{cases} \quad (7.7)$$

Assigning a value of zero in cases where $N_i^{(k)} \leq 1$ reflects the absence of observable variability. These zero values are therefore structurally present in the non-log-transformed dataset used for the LPA.

Initial Deposit

The initial deposit is defined as the amount recorded in the first deposit transaction. This serves as a proxy for the intensity at the moment of onboarding.

$$\text{InitialDeposit}_i = \begin{cases} x_{i(1)}^{dep}, & \text{if } N_i^{dep} \geq 1 \\ 0, & \text{if } N_i^{dep} = 0 \end{cases} \quad (7.8)$$

Here, $x_{i(1)}^{dep}$ represents the earliest observed deposit amount of user i ordered in time, and N_i^{dep} the total number of deposits of user i . Assigning a value of zero to users without deposit activity reflects the absence of observed entry funding.

¹Included: Bitcoin, Litecoin, Monero Tether USDt, USDT20, Monero, Ethereum, Dash, TRON, USDC, TrueUSD, Dai, PM and USD.

Dominant Asset Share

Here, E_{ic} represents the number of transactional events (deposits + withdrawals) of user i using currency c . The denominator $E_i = \sum_c E_{ic}$ calculates the total number of transactions.

$$S_i = \frac{\max_c(E_{ic})}{E_i} \quad (7.9)$$

This yields in a proportional value of transactions conducted in the user its most frequently used currency.

Number of Distinct Assets

A is equal to the number of unique currencies c used by user i . Where E_{ic} is similar to the previous equation and represents the number of transactional events by user i with currency type c .

$$A_i = |\{c : E_{ic} > 0\}| \quad (7.10)$$

Within this operationalization, transaction variability (7.7) and initial deposit behavior (7.8) were retained in their original amount units. As the primary interest of this study lies in identifying large deviations or consistency in behavior next to "entry" behavior, rather than explicitly quantifying the financial exposure, which would likely become visible through the value-based indicators already, regardless of currency conversion. Converting these indicators to USD-equivalents would introduce additional complexity and assumptions, such as the use of exchange rates and precise transaction-level price information, which are not required for this exploratory analysis and its behavioral focus.

7.1.2. Handling Missing Values and Structural Zeros

Due to the platform-specific data-generating processes, such as the presence of optional fields, but also the heterogeneity in platform usage and operationalization decisions, not all users engage in the same type of activities on the platform. Therefore, missing values are assumed not to be structural in nature. As a result, zero values are interpreted as the absence of specific behaviors or platform interactions, rather than being missing information, which would result in bias or requires imputation. This approach allows the model to capture distinct modes of financial platform usage rather than reflecting differences in overall activity intensity.

When reflecting at the operationalization, some zero values were intentionally created to reflect absence of behavior (e.g. for equations 7.1 or 7.2) or to ensure consistency when a variable could otherwise not be reliably computed (e.g. 7.5).

7.1.3. Indicator Distributions and Transformations

All indicators exhibit strong skewness and extreme values. As the analysis focuses on *relative* differences in behavior rather than exact scale interpretation, and to enhance model stability, we reduce the influence of outliers, by applying log-transformations² prior to standardization.

In the base model, all observations, including extreme values were retained. As the dataset represents raw empirical data, outliers may carry substantial behavioral information. To assess the sensitivity of the latent profile solutions to such extreme observations, two separate models were estimated: one including all observations and one in which the influence of extreme values was mitigated. The results of this sensitivity assessment is discussed in section 7.4.

After the log-transformation, all indicators were standardized by computing z-scores to ensure comparable scaling across variables prior to model estimation, illustrated by equation 7.11. This represents the final estimated base model.

²Log-transformations were implemented as $\ln(1 + x)$ for all variables to accommodate zero-valued observations.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (7.11)$$

7.1.4. Correlations

Prior to model estimation, bivariate correlations of the indicators of section 7.1 were inspected to assess potential redundancy and thus over representation of specific behavioral dimensions that would obscure qualitative differences in behavioral patterns. Moreover, as LPA assumes local independence, implying that within each class observed indicators are conditionally independent, it is worthwhile to assess correlations (Sinha et al., 2021).

Due to the strong skewness of the distributions in the applied dataset, Spearman correlations were calculated to assess bivariate correlations. Although there is no universally accepted critical threshold for problematic correlations defined in literature (Sinha et al., 2021), a threshold of $|\rho| > 0.7$ was applied for variables that need further inspection and potential removal (Dormann et al., 2013), together with assessment of the behavioral dimensions of the previous chapter. Therefore, this is applied in a diagnostic manner, rather than as a strict exclusion criterion. Only if it can be assumed they imply substantial conceptual redundancy and do not capture different behavioral dimensions (Chapter 5, they are removed. A more detailed analysis of the bivariate correlations is presented in Appendix X.

Based on this criterion, the indicators *account lifespan*, *activity days*, and *number of distinct assets*, *average withdrawal value (USD)*, *average deposit value (USD)*, *number of withdrawals* and *standard deviation of deposits* were removed. Among the remaining indicators, *number of deposits* serves as a broader measure of transaction frequency, and *standard deviation of withdrawals* as a broader measure of transaction size variability.

7.1.5. Final Conceptual Model

Based on the theoretical framework developed in Chapter 5, the inspection of distributions and scales, Figure 7.1 presents the conceptual model, which outlines the selected indicators to distinguish latent user profiles onto. For these indicator variables, their descriptives are also presented on their original scale in ???. Moreover, based on the applied log-transformations, box plots are presented in Figure 7.2, that indicate severe outliers still present in the dataset.

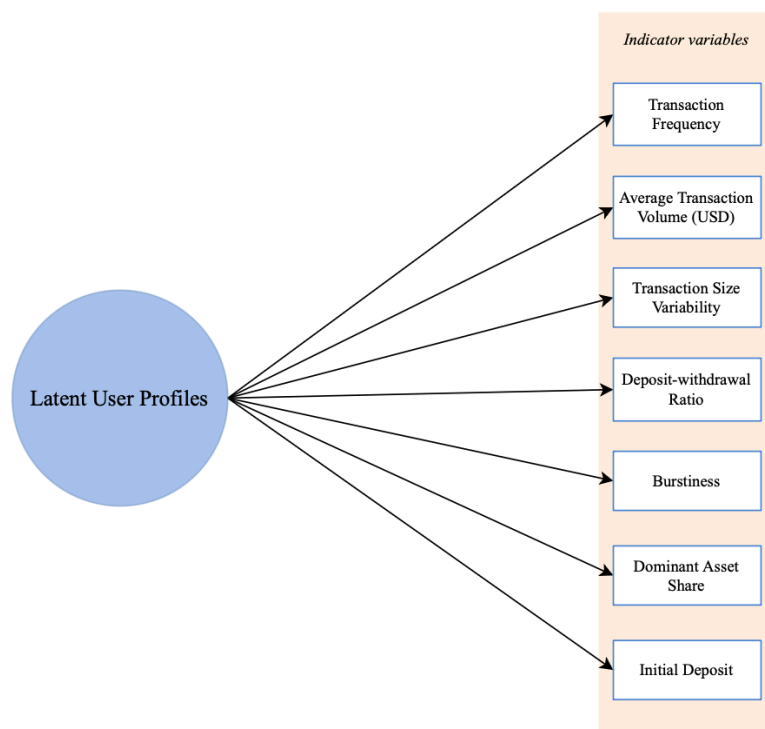


Figure 7.1: Conceptual model for the LPA

Table 7.2: Descriptive Statistics of Indicator Variables

Variable	Mean	Median	SD	Minimum	Maximum
Number of Deposits	660.51	3.00	8955.65	0.00	143 734.00
Average Transaction Volume (USD)	37 972.62	1157.78	112 393.19	0.00	1 256 083.70
Withdrawal Amount Variability	25 376.65	3.96	74 432.34	0.00	449 826.45
Deposit-to-Withdrawal Ratio	12.70	1.00	178.88	0.09	2874.70
Burstiness	0.15	0.09	0.25	−0.93	0.79
Dominant Asset Share	0.76	0.84	0.25	0.12	1.00
Initial Deposit Amount	18 825.75	99.50	83 488.99	0.00	800 000.00

Note. Only the average transaction size is expressed in USD.

Note. Number of Deposits reflects overall transaction frequency (deposits and withdrawals combined), and Withdrawal Amount Variability reflects overall transaction size variability across both transaction types. Original variable names are retained for accuracy.

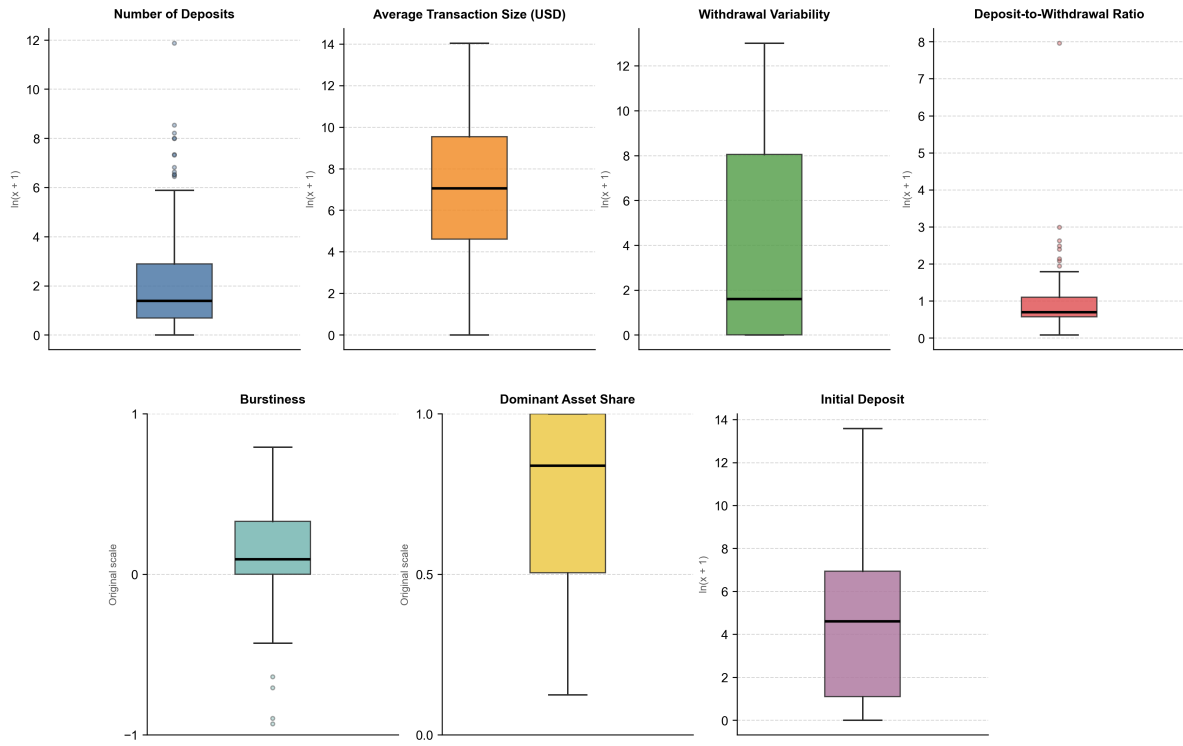


Figure 7.2: Distribution of indicator variables (log-transformed)

The descriptive statistics on the original scales reveal substantial heterogeneity across users, particularly for the monetary variables. The large discrepancies between means and medians. These are most notable for *initial deposit amount* and *average transaction size*, and indicate a strong right skewness, underscoring the necessity of log-transformation prior to analysis. The wide ranges observed across most indicators further suggest that users differ substantially in their engagement with the platform, providing a strong empirical basis for a person-centered analytical approach.

7.2. Results

7.2.1. Relaxation of the Local Independence Assumption

Initially, a maximum of ten cluster models, assuming local independence, were estimated. However, after this initial estimation inspection of bivariate residuals (BVRs) indicated substantial residual associations between pairs *standard deviation of withdrawals* and *average transaction value (USD)* and the *deposit-withdrawal-ratio* and the *number of deposits* as these BVRs were structurally high across all cluster models. The local independence assumption was relaxed for these specific pairs by allowing the residual covariances to be freely estimated, to prevent overlapping information from disproportionately influencing class assignment (Vermunt & Magidson, 2002). This adjustment is considered appropriate, as this decision is guided by substantive mathematical considerations, rather than that the correlations are assumed to be sample specific (Nylund-Gibson & Choi, 2018) ³. A more detailed discussion is provided in Appendix B.2.

The remaining maximum residuals per model concern only one or two indicator pairs and are considered tolerable, with all other BVRs falling below the critical threshold of 3.84 or only showing slight deviations⁴. These are interpreted as reflecting minor shared variance between conceptually related indicators, rather than meaningful structural misspecification.

³When indicators share substantial overlapping information, ignoring their residual association may lead the model to create artificial classes to absorb this shared variance. By allowing selected residual covariances, the risk is reduced that correlation is misinterpreted as substantive heterogeneity.

⁴This value of 3.84 corresponds to the critical value of a chi-square distribution with 1 degree of freedom at the 5% significance level. BVR values exceeding this threshold indicate statistically significant local dependence between indicator pairs

7.2.2. Model Fit Comparison

Based on the adjusted specification, models with one to ten classes were estimated and are presented in Table 7.3. The model is selected on the BIC and AIC, relative class size and substantive interpretability, as described in section 4.2 of the methodological chapter.

Table 7.3: Fit indices and stability diagnostics for latent profile solutions

Profiles	LL	AIC	BIC	Entropy	Max BVR	Min. class size (%)	LL Replicated
1	-2480.0751	4992.1502	5048.9976	1.0000	94.8036	-	Yes
2	-1844.3828	3754.7655	3872.0132	0.9923	26.1884	33.7	Yes
3	-1583.6988	3267.3976	3445.0456	0.9819	13.6928	23.0	Yes
4	-1401.3948	2936.7896	3174.8379	0.9644	8.4489	11.3	Yes
5	-1309.3374	2786.6748	3085.1234	0.9339	7.2525	3.9	No
6	-1194.8406	2591.6811	2950.5300	0.9670	6.0843	2.0	No
7	-1131.7109	2499.4217	2918.6710	0.9537	5.2106	2.6	No
8	-1064.7159	2399.4319	2879.0814	0.9646	5.3142	2.7	No
9	-1015.0188	2334.0376	2874.0875	0.9559	3.7892	2.0	No
10	-948.4212	2234.8425	2835.2926	0.9728	3.5362	2.5	No

Note. LL = log-likelihood; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion.

Additionally, since a consistent decrease in fit indices is visible, an elbow plot is presented Figure 7.3 for the various statistical indices, following a recommended practice by Nylund-Gibson and Choi (2018) and Sinha et al. (2021) to assess the optimal number of classes. Although the BIC and AIC continue to decrease, the relative reduction after four to five classes becomes considerably smaller. Visually, the four-class solution seems to offer the best fit based on the marginal improvement in the elbow plots of Figure 7.4. However, when balancing this with relative class sizes, the substantive investigative interpretability as well as statistical fit indices, the five-class solution was ultimately selected.

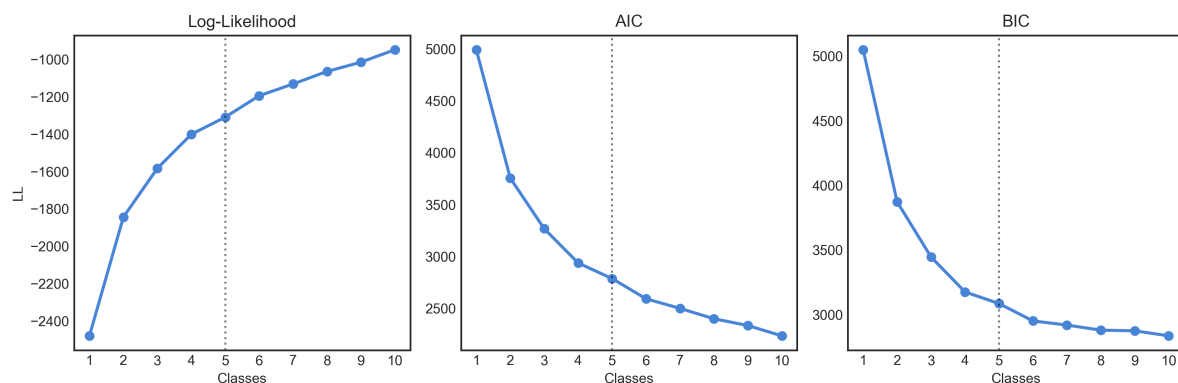


Figure 7.3: Statistical fit indices for the base model

Moreover, from a substantive viewpoint, as the number of profiles increases in Table 7.3, the additional segmentation is mainly driven by differences in overall activity levels. This effect is also known as the "salsa-effect" mentioned by Sinha et al. (2021). The five-class solution captures this differentiation meaningfully: the largest profile in the four-class solution with a size of 41.1% of the study population is split into two substantively distinguishable subprofiles, balancing out the overall class sizes. This distinction is partly quantitative in nature, indicating resemblance to the explained "salsa-effect". However, within this research differences in activity intensity are considered to represent relevant implications for investigative contexts, as different quantitative levels may reflect different motives or levels of criminal involvement. Additional cluster models start to split the more extreme observations specifically on the high activity dimension, rather than observing distinct behavioral patterns, creating very small clusters and model instability. Although the five-class solution includes a smallest class of 3.9%, which falls below the recommended minimum of 5% (Nylund-Gibson & Choi, 2018; Sinha et al., 2021), this is considered acceptable given the exploratory nature of the study and the substantive interpretability of the

profiles. However, beyond five classes, the cluster sizes decrease substantially, which increases the risk of overfitting and results in profiles that are difficult to interpret substantively. Profile solutions of the three-, four- and five-cluster models can be found in Appendix B.4 for reference.

Taken together the model selection criteria of subsection 7.2.1 and the elbow plot in Figure 7.3, this supports the retention of the five-class solution for the 'classify-analyze' approach within this thesis and its subsequent analyses.

7.3. Profile Overview

The five-class latent profile solution identified five distinct user segments on the exchange. Within this section, a distinction is made between (i) behavioral structure and (2) behavioral magnitude, which resembles the variation found in the indicator variables per profile. These descriptions represent empirically derived user segments that can help researchers and practitioners understand overlap and differences across the *behavioral structure* of individuals (Weller et al., 2020). This has potential implications for subsequent analyses and law enforcement practice. The relative differences in z-scores on the indicator variables are depicted in Figure 7.4 to support interpretation of this *behavioral structure*, and the median values for the original value of each cluster in Table 7.4 to indicate the actual *behavioral magnitude*. A better overview on inner profile distributions is presented in section C.1 of the Appendix. Importantly, still significant variation in behavioral magnitude is found between users of the same profile, this means that substantial heterogeneity in individual user behavior still exists within each profile.

For each profile, potential illicit explanations for the observed behavior are considered based on existing literature and the three stages of money laundering (Levi & Soudijn, 2020). However, it is important to note that this research does not aim to explain or attribute the observed behavior to illicit activity, but rather to explore broader behavioral patterns and relations to risk variables in Chapter 8.

Following the assessment of multicollinearity, in Appendix section B.5 the distribution of mean z-scores across the 5-profile solution for the removed indicator variables can be inspected. The results indicate that the identified profiles largely explain the variation in these variables, thereby supporting the decision to exclude them from the LPA due to their limited additional discriminatory value.

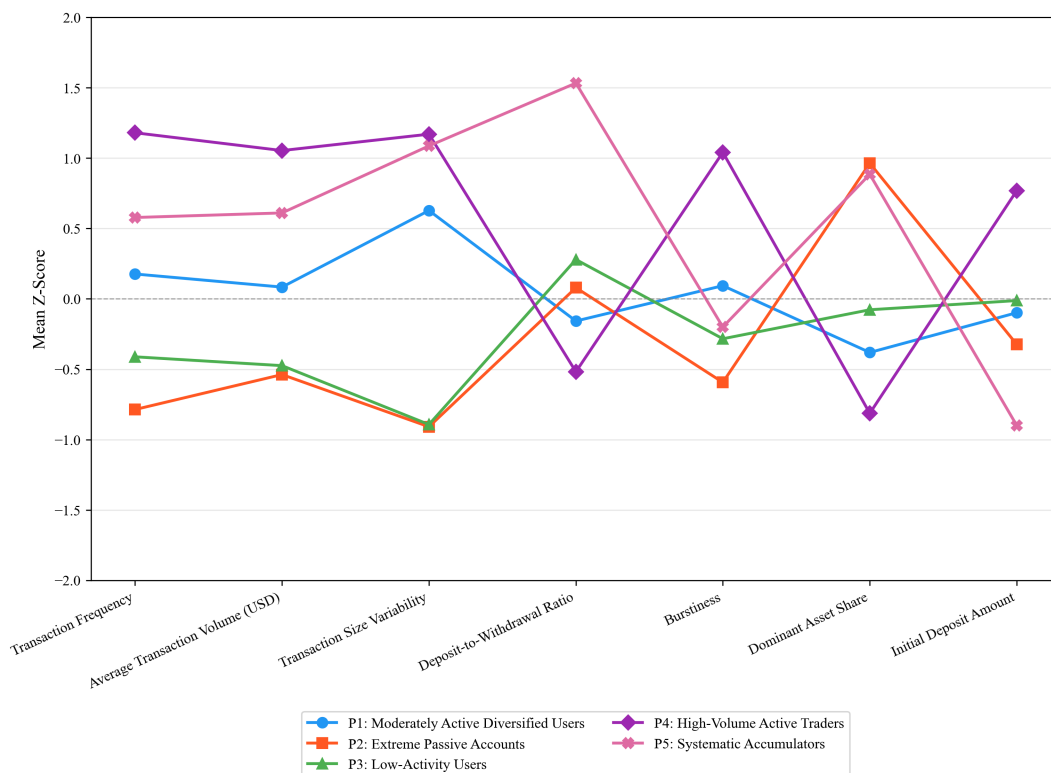


Figure 7.4: Mean z-scores for each cluster of the 5-profile solution

Table 7.4: Median indicator values per latent profile (original scale)

Indicator	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
Transaction Frequency	20.82	1.00	3.39	521.06	16,008.11
Average Transaction Size (USD)	14,908.00	22,426.46	22,772.52	116,439.27	38,347.79
Transaction Size Variability	17,363.21	0.01	0.13	81,434.72	155,570.27
Deposit-to-Withdrawal Ratio	1.24	1.60	2.18	0.69	323.58
Initial Deposit	4,770.04	1,053.66	10,775.07	80,597.84	10.89
Burstiness	0.18	-0.00	0.08	0.42	0.09
Dominant Asset Share	0.66	1.00	0.74	0.56	0.98

Note. Due to right-skewed distributions, only medians are reported. Additional boxplots are provided in section C.1 to offer further insight into the distributions of variables within the profiles.

7.3.1. Profile 1 (28.0%): Moderately Active, Diversified Users

Within this profile, users show relatively moderate levels of transactional activity. The transaction frequency and the average transaction size are close to the sample mean. The variability in transactions is relatively elevated, suggesting occasional larger and smaller transactions alongside more routine activity. Notably, this profile scores low on dominant asset share, indicating that activity is spread across multiple assets rather than concentrated in a single one. Burstiness and deposit-to-withdrawal ratio are both close to the sample average as well, which suggests relatively steady and balanced platform engagement. Overall, this profile resembles a moderate version of the high-activity trading behavior in Profile 4. Users are engaged with the platform across multiple assets, but without the extreme intensity observed in the fourth profile. From a law enforcement perspective, the diversified asset usage combined with a relatively higher variable transaction size behavior may be consistent with layering activity, though at a less noticeable level than the most active profile.

7.3.2. Profile 2 (25.6%): Extreme Passive Accounts

This profile is characterized by representing the lowest transactional activity. The transaction frequency, average transaction size, and withdrawal variability are all substantially lower than the sample mean, which also explains why the dominant asset share is the highest across all profiles, as well as the burstiness score. Based on this, the users seem to be concentrated passive holders of assets, potentially suggesting a buy-and-hold or try-out strategy rather than active trading on the platform. This could potentially be consistent with placement behavior, parking illicit funds in a stable asset while avoiding attention through inactivity, but generally, funds are often moved faster to limit traceability. But before drawing any conclusions on this observation, this will be further explored, discussed and validated with experts from law enforcement in Chapter 9.

7.3.3. Profile 3 (23.3%): Low-Activity/Try-Out Users

Users in this profile display the lowest levels of engagement across virtually all behavioral indicators. Deposit frequency, average transaction size, and withdrawal variability are all substantially below the sample mean. The deposit-to-withdrawal ratio is slightly positive, though this probably reflects the infrequency of the behavior. Overall, this group shows minimal and sporadic platform interaction, with no clear directional pattern in terms of asset concentration or transactional timing. The initial deposit is almost average, but generally all users seem to deposit a small amount first. Potentially, this implies that this profile may represent users who created an account but never fully engaged with the platform, or who became inactive after an initial period of limited activity. This can possibly indicate try-out behavior of the platform. While this inactive profile represents the least suspicion based on this perspective from a law enforcement perspective, this remains an important topic for discussion in Chapter 9.

7.3.4. Profile 4 (19.2%): High-Volume Active Traders

This profile includes users who are the most active and extreme segment of the study population. Within this profile, users score highest across nearly all activity-related indicators. The transaction frequency, average transaction size, withdrawal variability, and burstiness are all strongly higher than the sample mean. Because of the burstiness value, the activity tends to occur in concentrated bursts

rather than spread regularly over time. Similarly to Profile 1, this profile scores low on dominant asset share, indicating that activity is spread across multiple assets. The combination of high frequency, large and variable transaction amounts, diversified asset usage and the fact that withdrawals are also more frequent compared to their deposits among this group, it is potentially consistent with speculative trading behavior, or layering tactics. From a law enforcement perspective, this profile can be interesting as funds are moved frequently, in variable amounts, and across multiple assets, making the transaction trail difficult to follow.

7.3.5. Profile 5 (3.9%): Systematic Accumulators

Even though this profile contains less than 4% of the study population, Profile 5 is arguably the most behaviorally distinctive. Users in this group show a very high deposit-to-withdrawal ratio, by far the highest across all profiles, which indicates that substantially more funds flow in than out of the platform. Dominant asset share is also high, suggesting users within this cluster concentrate in a single asset. The burstiness is negative and thus lower than average, indicating more routine likely behavior. Initial deposit is the lowest across all profiles, suggesting that users in this group started with relatively small amounts before accumulating over time. The combination of structured timing, concentrated asset usage, and persistent accumulation of funds with minimal outflow sets this profile apart from the others. It potentially makes sense for these accounts to represent hot wallets and thus internal users on the platform.

7.4. Sensitivity Analysis of the Profile Solution

Evaluating the Impact of Outliers

As the base model includes all user-level observations present in the *core_deposit* and *core_withdrawal* datasets, a sensitivity analysis was conducted to assess robustness of the model outcomes towards outliers present in the dataset. As described earlier, these outliers were intentionally retained in the base model, due to the empirical nature of the dataset and the aim to preserve any rare but potentially relevant behaviors. To further ensure the stability of the results, 500 random starting points were used during estimation.

Outliers were defined by using the interquartile range (IQR) criterion ($1.5 \times \text{IQR}$) for each included indicator variable of Figure 7.1 on their log-transformed scale. This corresponds with visible outliers of Figure 7.2. Observations exceeding the IQR-boundaries on at least one indicator were excluded, after which the LPA was re-estimated and compared to the base model. As a result, 24 *user_ids* were considered as outliers, resulting in removal of 9.3% of the base model for this the sensitivity analysis. Below Table 7.5 depicts the number of outlier observations per indicator.

Table 7.5: Observations identified as outliers per indicator variable

Indicator variable	Outlier observations	Percentage
Number of deposits	12	4.7%
Deposit-withdrawal ratio	8	3.1%
Burstiness	4	1.6%

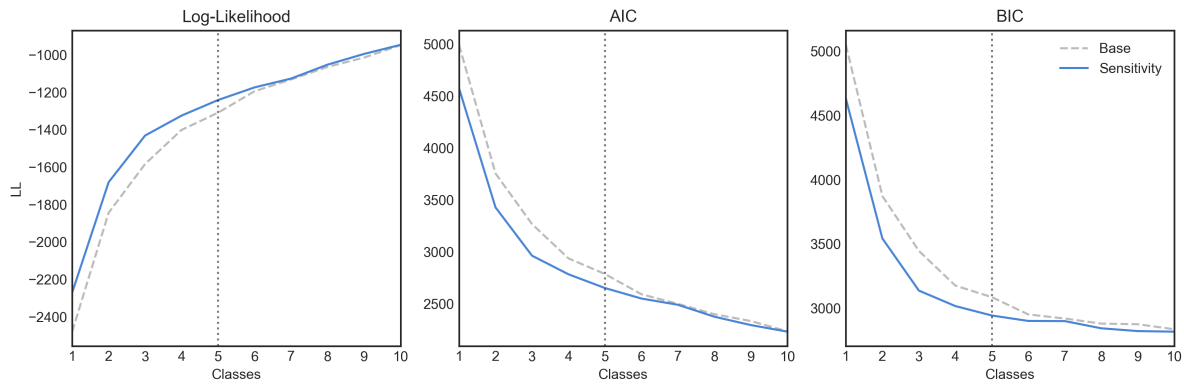


Figure 7.5: Comparison of model fit indices across latent class solutions: Base model and sensitivity analysis

The sensitivity analysis of Figure 7.5 without outliers shows a that pattern is consistent with the base model. Although the overall fit indices improved in absolute terms, as expected due to the reduction of noise (Appendix subsection B.6.1, the shape of the graphs and the relative improvement for each class solutions remains comparable. In both specifications, the largest improvement for the BIC and AIC occurred up to three to four to five classes, after which the decrease flattened substantially. The curves representing the sensitivity analysis in Figure 7.5 provide even stronger support for the five-profile solution.

Moreover, the results of the sensitivity model (Appendix B.6) recover the same five profiles from a substantive perspective, and the relative sizes of each class remains comparable overall (37.0%, 27.9%, 21.3%, 9.9%, and 3.9% versus 28.0%, 25.6%, 23.3%, and 19.2%, and 3.9% in the original model). The outliers do not form a separate cluster on their own. Across all clusters, the direction and pattern of the behavioral indicators remain similar, indicating stability. Mainly the scale of the standardized values is increased without inclusion of outliers, as variance is decreased by the exclusion of outliers, making the profile specific means more extreme. This suggests that the five-cluster solution reflects genuine behavioral patterns, rather than being influenced by extreme observations.

Importantly, this difference in scale mainly has implications for interpretation. When outliers are excluded, clusters represent more clearly separated behavioral segments, making them easier to interpret. In contrast, when outliers are included, clusters may combine typical observations with more extreme cases, which reduces the clarity in the distinction between the profiles, while the behavioral structure remains preserved.

All in all, the conducted sensitivity analysis supports the robustness of the five-class model. As the same behavioral patterns emerge, this increases confidence that differences across profiles are considered meaningful. Outliers do not change what the clusters are, but mainly how clearly they can be interpreted.

Evaluating the Impact of Internal Users

As described in subsection 6.2.4, the dataset includes multiple user accounts that are presumably related to Cryptex its internal administration. In total, based on e-mail addresses, 16 user accounts have been marked as internal based on their e-mail address structures in the study population. This implies that 242 records were retained for a separate sensitivity analysis.

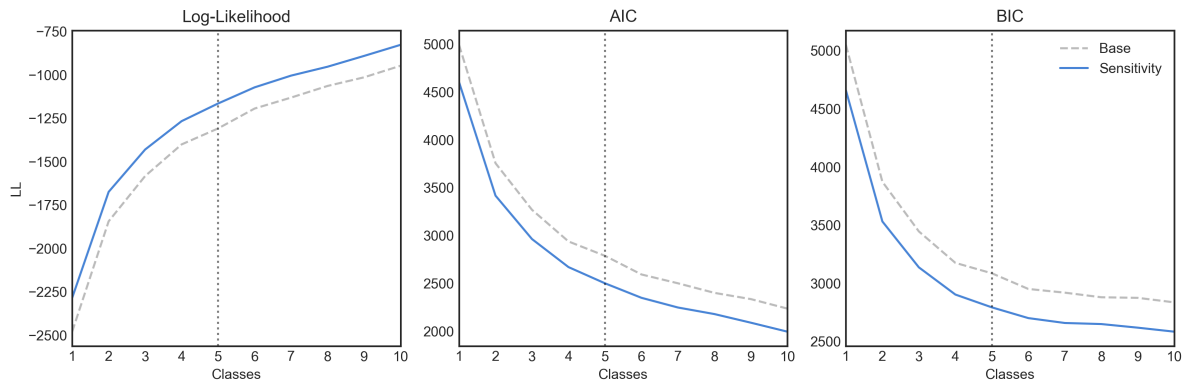


Figure 7.6: Comparison of model fit indices across latent class solutions: sensitivity analysis without internal users

Excluding internal users from the LPA is found not to materially affect the model selection criteria. The shape of the AIC and BIC curves remains consistent, even more so than the plots without outliers, with the elbow visible around five classes. This further supports that the identified cluster solution is not driven by administrative accounts, but reflects genuine behavioral patterns in the user population.

More substantially, which can be derived in more detail from subsection B.6.2 in the Appendix, the three behavioral dimensions activity, variability and asset concentration remain dominant for differentiation and the cluster structure remains largely consistent. Moreover, the fifth cluster remains, although smaller in absolute size. Overall, excluding the internal users from the LPA mainly refines the profiles instead of altering the underlying structure. This supports the robustness of the initial model, and indicates that the behavioral variation is driven by user-level, instead of platform-level differences.

7.5. Interpretation Summary and Implications

The results show that activity intensity is a dominant dimension in structuring the latent profiles. Across all model solutions, users are primarily clustered based on transaction frequency and variability. Asset concentrated is a second, clearly distinguishable dimension, that separates diversified users from users that focus on a single asset. Finally, capital-flow dynamics, in this case the deposit-withdrawal ratio provides a third, more nuanced layer in refining the profiles.

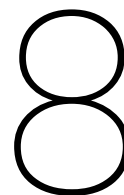
These findings show that segmentation is driven by behavioral dimensions. It is important to note that these dimensions reflect relative differences between users rather than absolute levels of activity, highlighting the important difference between behavioral structure and magnitude. As a result, users within the same profile can exhibit similar behavioral patterns while these may occur at substantially different scales in terms of transaction volume and value, due to the log-normalization and standardization. At the same time, the increasing number of classes introduces greater behavioral nuance but comes at the expense of model stability. Higher-class solutions tend to fracture extreme observations into smaller, less stable groups. This complicates generalizability across samples and model runs.

Moreover, as LPA operates on continuous variables and assigning them to a finite number of profiles, it implies the existence of qualitatively distinct subgroups within the population. However, it can not be guaranteed that the identified profiles represent theoretically or empirically grounded groups. Instead, it can be merely a reflection of the model its tendency to divide the continuous distributions into more detailed clusters that capture minor variations rather than distinct behavioral patterns. Therefore, the profiles have not been provided concise names, but rather a summarizing description on their behavior, to avoid concluding the presence of specific profiles.

This trade-off highlights a broader methodological tension. LPA is fundamentally designed to uncover stable population structures. However, in investigative contexts, extreme or anomalous behaviors are often of particular interest, also shown by previous research of e.g., Gu et al. (2022), Hu et al. (2025), and Jiang et al. (2022) in Chapter 1. However, these cannot be detected without a sufficient overall sample size to estimate classes of relatively low prevalences (Nylund-Gibson & Choi, 2018). Another issue contributing to this is the fact that extreme users may not necessarily form distinct behavioral profiles, but instead represent high-magnitude instances within otherwise similar behavioral

patterns. This limits the ability of LPA to provide clear guidance on the most 'true' segmentation in such research contexts. As a result, the five-profile solution reflects a deliberate choice for this research: balancing statistical marginal gain with interpretability over maximal behavioral differentiation. While higher-class solutions offered additional nuance, they mainly captured extreme observations and reduced replicability, which completely loses the principle of LPA as a method.

A final consideration involves the continuous decline of the BIC across model solutions in Table 7.3 and Figure 7.3, without reaching a clear minimum. According to Nylund-Gibson and Choi (2018), this may indicate that the mixture model is not correctly specified, and that alternative model specifications could potentially provide a better fit. However, given the exploratory nature of this study and the strong substantive interpretability of the four-profile solution, this limitation is acknowledged without adjusting the model specification. The five-profile solution is therefore retained as the primary model.



Post-Hoc Profile Characterization

Following the Latent Profile Analysis (LPA) presented in the previous chapter, this chapter further contextualizes the identified behavioral profiles through several post-hoc analyses. For this analysis, the profiles are evaluated based on whether they exhibit consistent behavioral deviations across (i) network structure, derived from address-sharing relationships, (ii) blockchain illicit exposure indicators and (iii) indicators related to internal KYC-compliance procedures. The underlying rationale is to assess whether the identified profiles differ beyond the variables directly included in the LPA and whether profiles that show converging signals that could be considered operationally or investigatively relevant. Conversely, profiles for which such signals are absent or inconsistent may reflect more heterogeneous behavior or the use of alternative transaction mechanisms that are not captured within the current lens through which the behavior is analyzed. As such, this chapter aims to further refine the interpretation of the profiles as well as critically assess whether the identified latent profiles represent substantively distinct patterns beyond the transactional variables within Cryptex.

By doing so, this chapter contributes to answering SQ3: *“How do the identified user profiles differ in behavioral patterns associated with increased risk?”* and SQ4: *“What insights can be derived from the identified profiles and their associated risks patterns?”*

8.1. Analytical approach

For this analysis, no specific temporal ordering or causal relationships are assumed between class membership and the analyzed variables. This chapter therefore focuses on the descriptive associations between the profiles and these additional variables, externally from the LPA framework.

As described in Chapter 4, a classify-analyze procedure is followed. This is one of the procedures for assessing distal outcomes in LPAs (Nylund-Gibson et al., 2019). The five-profile solution showed a considerable high entropy (0.9339), with the majority of observations showing substantially higher certainty of classification (posterior probability > 0.80). This supports the use of a hard classify-analyze procedure as outlined in Chapter 4 (Clark & Muthén, 2009). However, it should be noted that the aggregated entropy value can hide classification uncertainty at the level of individual profiles, which introduces potential misclassification bias. This represents a known limitation of this approach (Nylund-Gibson et al., 2019). To examine classification certainty at the profile level, the mean maximum posterior probability and misclassification rate were computed for each profile and are presented, as well as the proportion of observations assigned to each profile with a maximum posterior probability below < 0.80 in Table 8.1. Compared to other profiles, the first and fourth cluster display a notably higher misclassification rate of 19.7%, with 10.6% of its observations falling below a posterior probability of 0.8. Findings within this post-hoc analysis using this “classify-analyze” method should therefore be interpreted with appropriate caution. This is particularly relevant when interpreting profile-specific post-hoc differences, since profiles with a higher classification uncertainty may have more behavioral overlap with adjacent profiles.

Table 8.1: Classification certainty by latent profile

Profile	N	Mean Max. Posterior	Misclassification Rate	% Prob. < 0.80
Profile 1	76	0.9110	0.0890	19.7%
Profile 2	67	0.9879	0.0121	3.0%
Profile 3	59	0.9946	0.0156	0.0%
Profile 4	47	0.9455	0.0545	10.6%
Profile 5	9	0.9551	0.0449	0.0%

Note. Misclassification rate = 1 – mean maximum posterior probability.

As mentioned in Chapter 4, the results are interpreted by evaluating the observed profile-specific proportions with the overall proportion observed in the dataset. This assesses whether a characteristic is either over- or underrepresented within a profile compared to the average user in the overall study population. Deviations were expressed as percentage-point differences (Δ), where:

$$\Delta = P_{profile} - P_{overall}$$

Here, $P_{profile}$ is the percentage of users within a profile that is associated with the specific characteristic of the post-hoc analysis. $P_{overall}$ is the percentage over all profiles, thus the proportion observed in the full study population. Positive values indicate that a characteristic is more prevalent within a profile than in the overall sample, whereas negative values indicate lower prevalence. This descriptive measure was used to identify whether the variable under analysis is relatively over- or underrepresented across the latent profiles.

8.2. Profiles and Address Sharing Patterns

The network analysis was specifically intended to examine whether the latent behavioral profiles identified from internal exchange activity also exhibit distinct structural patterns within blockchain-related address-sharing behavior. As described earlier in the methodology of Chapter 4, the analysis focuses on separate construction of deposit-based and withdrawal-based undirected networks in which links represent use of an identical withdrawal address. This shares characteristics with the definition of a social network (Rabade et al., 2014). However, here an important difference is that the network is based on similarity in addresses, not direct (transactional) interaction. In this analysis, the specific type of deposit or withdrawal is not taken into account, which limits visibility in similar off-ramp or cash-out strategies, or address overlap for cash-drop offs for example. Therefore, it is considerable that only overlap is found in blockchain wallet addresses.

Within the network, a link between two users is established when they share an identical value in the deposit or withdrawal address field. Here, empty fields were excluded, as well as entries such as "test", or single-character values to clean the network. However, it should be noted that these values reflect user-provided input and therefore do not necessarily have to correspond to actual blockchain addresses. Entries including variations of "test" were mainly found in the deposit network, in which 62 records were removed, compared to 9 in the constructed withdrawal network. Within the network, edge weights represent the number of unique withdrawal wallets shared by each pair of users.

Across both network types, connectivity remains relatively limited after removing internal accounts, as the deposit address network includes 4 nodes (users) and 4 edges (shared wallet addresses), while the withdrawal network consists of 41 nodes and 58 edges as presented in Table 8.2. This reflects that shared address usage occurs only among a subset of users and the presence of internal users significantly contributes to these observed patterns.

The analysis was conducted excluding the internal accounts, but have been accounted for in Table 8.2. When internal accounts are included, the network becomes substantially more connected with a larger number of edges and a more extensive connected component as can be derived from Table 8.2, but it is assumed that this potentially creates artificial connectivity. After excluding these accounts, the network becomes more fragmented, with fewer connections and a higher number of isolated users as presented in Table 8.2.

Table 8.2: Network overview per constructed network

Metric	Deposit	Deposit (no internal)	Withdrawal	Withdrawal (no internal)
Number of nodes	9	4	62	41
Number of edges	10	4	125	58
Clustering coefficient (<i>weighted</i>)	0.14	0.37	0.047	0.028
Clustering coefficient (<i>unweighted</i>)	0.22	0.58	0.41	0.34
Number of connected components	2	1	4	5
Largest connected component size	8	4	56	33

Notably in Table 8.2, the majority of users do not appear in either of the networks, implying they do not share any wallet address with other users. This absence from the network is itself an informative characteristic. Since both deposit networks are substantially small, no network analysis is applied and these networks are neglected for further interpretation. It can be concluded however, that address-sharing patterns across users mainly happens when moving funds from the platform to other destinations.

When considering the global networks in Table 8.2, it becomes clear that the weighted clustering coefficient is lower than the unweighted variant. This suggests that many triadic relations are supported by relatively weak overlap ties. This means that shared address usage not only mainly happens when withdrawing funds, but also tends to occur sporadically rather than indicating repeated overlap between the same users.

Overall, the withdrawal network appears to consist of one dominant connected structure alongside several smaller, isolated clusters. This suggests that relatively a large subset of users is embedded in broader shared-wallet overlap patterns, while some users are fully disconnected or are only connected within small localized clusters. As a result, the network reflects a combination of widespread infrastructural overlap and more isolated address-sharing behavior. Moreover, as mentioned, it can be concluded that the internal users are generally responsible for connecting users, because the number of nodes significantly increases for both network types that include these accounts. The network is further assessed at the node level to examine structural differences in network position between the profiles

8.2.1. Interpretation and Limitations of the Withdrawal Network

As further analysis specifically focuses on the withdrawal network, it is important to distill meaning behind the network for interpretation purposes, as well as for choosing appropriate network metrics. Compared to the deposit network, withdrawals are more likely to represent deliberate or potentially strategic behavior by users, as it includes active decisions on where funds are transferred to after usage of the platform.

Taking this into account, the withdrawal network can provide two important types of insights: it may reveal information on individual (strategic) user behavior, for instance through patterns of address reuse or selective withdrawal destinations. Second, it can highlight user accounts or structures that can be relevant for further investigation, or it may offer insights into the broader transaction and batching processes within Cryptex.

Within this context, an edge represents shared withdrawal address information between user accounts, which may have multiple underlying explanations:

- **Common ownership:** multiple accounts controlled by the same entity, reflected in consistent reuse of withdrawal addresses.
- **Shared service usage:** users may withdraw to the same external services, such as exchanges, custodial wallets, or cryptocurrency mixers. This can result in indirect connections visible in the constructed network. However, as mentioned before, these connections do not have to imply a direct relationship between two users.
- **Coordinated behavior:** accounts that share withdrawal addresses may imply strategic coordinated behavior to transfer funds to the same wallet

- **Administrative or operational processes:** platform-related mechanisms (e.g., batching or internal wallet structures) that generate recurring address overlaps.

The network metrics presented and chosen for analysis in the following subsection 8.2.2 may provide an indication on either one of these explanations.

However, when interpreting the initial network properties, it is important to acknowledge limitations in examining address-sharing overlap approach. First, it only captures similarities in blockchain withdrawal behavior and is therefore unable to represent other operational strategies or cash-out methods known to exist on the platform, such as cash drop-offs or the use of Cryptex vouchers. This explains why only a subset of users is largely visible. As a result, the network reflects overlap in blockchain addresses rather than direct transactional interactions or actual fund flows between users. Consequently, coordinated behavior can only be identified when users share or reuse withdrawal addresses, potentially underrepresenting other ways of coordination that occur outside the observable blockchain withdrawal infrastructure. Users relying on alternative cash-out strategies, privacy-enhancing services, or off-chain transactions may therefore remain structurally invisible within the constructed network.

Importantly, the obtained network should therefore not be interpreted as a representation of the complete transactional system, but rather as a partial projection of shared withdrawal behavior within the observable *Bitcoin*, *Ethereum* and *Tron* infrastructure. Apart from the exposure-related indicators of section 8.3, the current approach provides limited insight into the broader positioning of users within the wider blockchain ecosystem. Directions for improvement of similar analyses are provided in the Discussion of Chapter 10.

8.2.2. Network Metrics for Analysis

For the withdrawal network, the larger number of users present allows a more meaningful analysis, while taking into consideration that the network is relatively small and does not include all users present in the sample used for the LPA. As we are mostly interested in information on the user level to evaluate differences between profiles, the focus is put on local centrality metrics (Scott, 2000). From an investigative perspective, and following arguments for choosing network metrics discussed by Morrison et al. (2022), main interest lies in the extent of *connectivity* between users, the *strength* of those connections and the *position* within the network. After visual inspection of the obtained network (see Appendix D.1), these aspects were aimed to capture by choosing the following network centrality metrics were for further analysis (Golbeck, 2013):

- **Degree centrality:** Degree centrality measures the number of nodes an individual node is connected with (Rabade et al., 2014). In this case, with how many individuals a user shares an address with. It can separate highly connected users, that potentially use services from users that operate more separately. As a result, a higher degree may imply interaction with external services that are frequently used or shared infrastructures, while a low degree may reflect more isolated behavior or limited address reuse. However, this unweighted degree centrality is only reported in Appendix D.1, as it is mainly included as a baseline measure of network connectivity. Because the network specifically captures varying levels of withdrawal-address overlap, the main analysis focuses on weighted network measures explained below.
- **Weighted degree centrality (strength):** Complementary to the degree centrality, the weighted degree (here defined as *strength*) is included. In this study, strength reflects the total amount of shared withdrawal-wallet overlap a user has across all connected users (Candeloro et al., 2016).
- **Weighted betweenness centrality:** The weighted betweenness centrality provides information on the connectivity of a user in the network by assessing how often it is part of a (weighted) shortest path between two users (Rabade et al., 2014). Here, this measures how often a user occupies structurally intermediary positions between different shared-wallet overlap structures. Comparing to other network types, here, the applied metric does not reflect transaction flows in this network, but rather the extent to which users connect different groups or cash-out strategies through shared withdrawal addresses, while taking the strength of these connections into account. Unweighted betweenness treats every overlap relation equally, regardless of whether users share one or many wallets. Weighted betweenness additionally accounts for overlap strength, which was more appropriate for this study because the network was specifically based on varying levels of shared withdrawal-wallet overlap. Here specifically, a high weighted betweenness centrality

score may indicate that a user occupies a structurally intermediary position between different shared-wallet overlap structures. This suggests that the user is connected across multiple parts of the network through relatively strong overlap relations, rather than being confined to a single local cluster.

- **Component presence:** Lastly, although not considered a centrality metric, the size of the component the user is present in is considered a meaningful characteristic. In the context of this analysis, smaller, detached components are considered to provide more meaningful information as they are less likely to reflect widespread service usage or highly connected infrastructural nodes, but may rather imply strategic behavior or common ownership.

8.2.3. Withdrawal Network: Results

For the network analysis, internal or administrative accounts were excluded, resulting in a remaining sample of 243 users (also visualized in Figure 8.1). As a result, the distribution of users across profiles differs slightly from the original LPA profile distribution. These differences are limited and are taken into account when interpreting the network results. This mainly has an influence on profiles 1, 4 and 5, where the internal accounts respectively constitute 7.9%, 10.6% and 11.1% of the cluster sizes (see Appendix C.2). Since Profile 5 is relatively small, the observed difference can also be attributed to random variation.

Figure 8.1 depicts the presence of users in the withdrawal address-sharing network per profile. A graphical version of network can be found in Appendix D.1. As also shown in Table 8.2, 41 users are present in this network, which is 17% of the population without internal users.

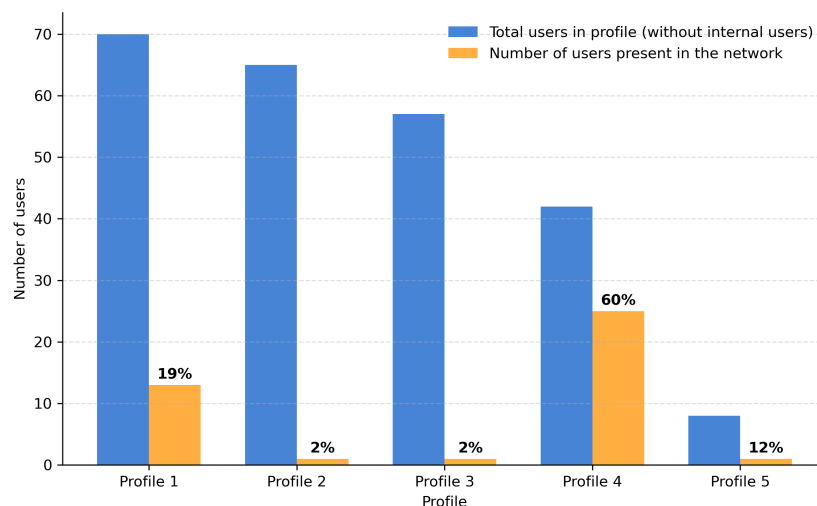


Figure 8.1: Proportion of users present in the withdrawal-based address-sharing network across profiles

Figure 8.1 shows that the extent of address-sharing patterns is highly uneven across the profiles. Profile 4 shows a significantly higher presence in the network. This suggests that address-sharing behavior is primarily associated with this cluster, while other profiles seem to operate more independently. Second comes Profile 1, with 19% of its non-internal users present in the withdrawal-based network. Due to the limited number of users present in the network for Profiles 2, 3, and 5 (only 1 observation per profile), meaningful distributional analysis is not considered feasible for these profiles. Therefore, the analysis of network metrics focuses on 1 and 4, where it is considered that sufficient observations are available to gain further insight.

When assessing the centrality network metrics discussed in subsection 8.2.2, depicted in Figure 8.2, it becomes clear that users in Profile 4 overall have significantly higher and varying values in strength and weighted betweenness centrality, indicating connections to a larger number of other accounts, also with a higher frequency of using common withdrawal addresses. This suggests broader and repeated engagement with shared address patterns, potentially reflecting interaction with widely used services or

infrastructure. On the other hand, users in Profile 1 show lower strength values, which indicates more incidental use of shared infrastructural patterns and a less central position within the internal network, this is supported by the percentages of presence in the largest component per profile, which was 69% for Profile 1 and 92% for Profile 4. Such small clusters may be considered more notable, as they could reflect very specific address sharing between specific accounts, which may reflect more selective or repeated use of shared withdrawal infrastructure. Potentially, this could indicate common ownership or coordinated transactional behavior in some cases. At the same time, a large proportion of users that mainly belong to Profile 4, are located within the single large component which consists of 33 nodes and shows more similarities in wallet-addresses. This suggests that users in this profile are more structurally embedded in shared address patterns, compared to Profile 1. A possible interpretation is that these users make frequent use of external blockchain services, such as exchanges, which may lead to repeated overlap in addresses. This may also be partly explained by their general elevated activity level within Cryptex, which can be associated with increased activity on the blockchain. Alternatively, this pattern may reflect a broader cluster of users exhibiting similar and interconnected behavior.

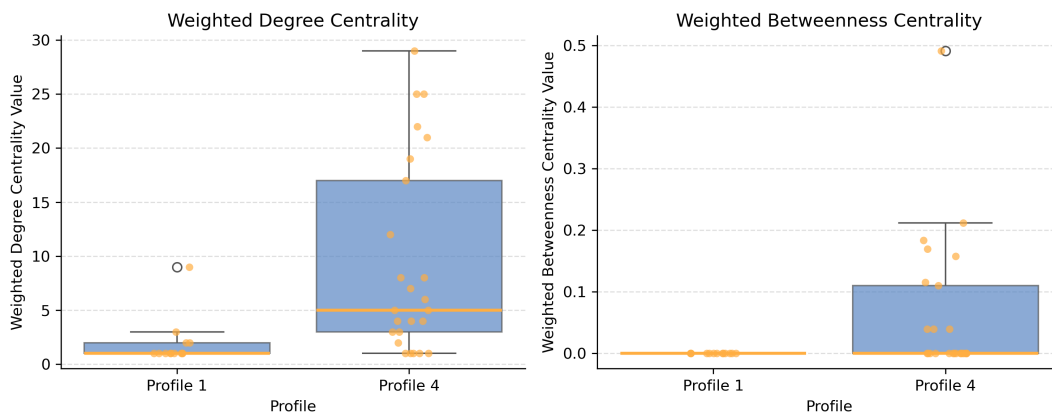


Figure 8.2: Distribution of centrality measures for Profiles 1 and 4

While Profiles 2 and 3 are known to show substantially lower activity levels overall, this alone does not fully account for the near absence of address sharing, as many of these users in the profiles do perform a few withdrawals. This suggests that users in these profiles may follow different and less interconnected transaction patterns, which may be due to alternative off-ramp strategies, not associated with a cryptocurrency blockchain.

Overall, although not statistically tested, the results suggest substantial differences between the profiles in their withdrawal-related address-sharing behavior. Specifically, Profile 4 is most prominently represented within the network and also with the highest repeated address-sharing behavior, indicating stronger involvement in shared withdrawal infrastructures or widely used external services. However, the lower-activity profiles are largely absent from the network. This suggests that users within these groups potentially rely on different transactional or off-ramp strategies that remain less visible within blockchain-derived overlap structures. As a result, it can be derived that observable address-sharing patterns capture only a subset of the broader transactional behavior present within the exchange environment.

However, the observed network differences should be interpreted by taking the behavioral characteristics identified in the LPA into account. This is because Profile 4 also exhibited the highest overall transactional activity, which likely contributes to the elevated connectivity and strength observed within the withdrawal network. Similarly, Profile 1 showed relatively more moderate activity levels and a comparatively higher degree of network presence than the other profiles. The results demonstrate nonetheless that the identified behavioral profiles differ not only in transactional behavior at the exchange level, but also in their structural positioning within the constructed address-sharing network. This suggests that the behavioral distinctions identified through the LPA may extend to broader observable blockchain-related withdrawal patterns. Although these findings should not be interpreted as evidence of illicit activity, the address-sharing behavior observed particularly within Profiles 1 and 4 may still be relevant

from a law enforcement or investigative perspective, as repeated overlap in withdrawal infrastructure can be associated with more complex transactional structures, shared service usage, or coordinated operational behavior in some cases.

8.3. Profiles and Receiving Exposure

Complementary to the network analysis, a blockchain illicit exposure analysis was conducted, by matching deposit and withdrawal addresses with associated risk labels provided by Chainalysis (Chainalysis, 2019a).

As explained in Chapter 4, only direct receiving exposure was considered to reduce false-positive characterizations of the obtained profiles. Moreover, this risk was operationalized as a binary variable, without distinguishing between specific risk categories. These indicators should be interpreted as risk signals embedded in transaction networks, not as definitive classifications of user intent or behavior.

Figure 8.3 shows the percentage of users with at least one external addresses used for deposits associated with illicit exposure.

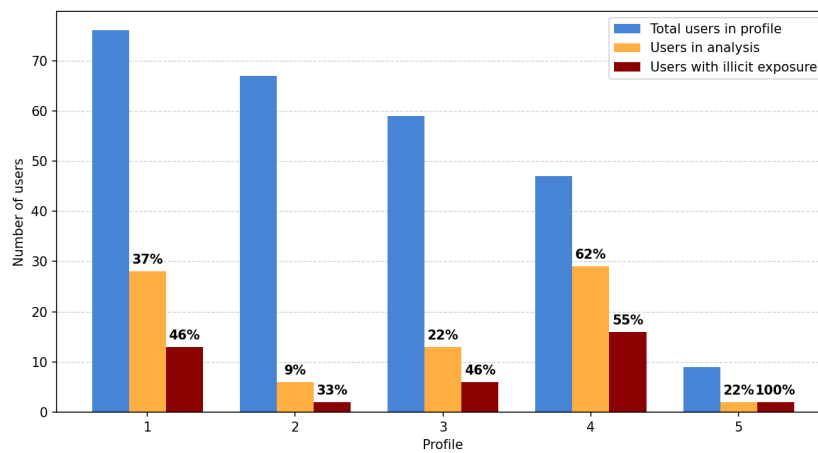


Figure 8.3: User distribution and illicit exposure on deposit addresses per profile

Note. Orange bars represent users using actual blockchain addresses. Red bars indicate users with illicit exposure. Percentages above bars are calculated relative to the total number of users within that profile.

A first observation is that a substantial share of the 234 unique users in the deposit dataset is not represented in this figure, as only 78 users are captured (sum of the blue bars). This indicates that only a part of all users actually provides valid blockchain addresses from which deposits are made. This discrepancy suggests the use of alternative deposit methods, but also reflects the presence of non-informative entries, such as test inputs and empty fields, which inflate the overall dataset while not being actual completed transactions.

Moreover, it can be observed that Profiles 1 and 4 make the most use of valid blockchain addresses (visualized in the orange bars). Notably, these profiles also have relatively higher levels of illicit exposure. This indicates that the users that most actively engage on the platform, also have more on-chain related deposits, and are also more frequently associated with illicit fund flows. From an investigative perspective, this suggests that these profiles may be of particular interest, as they combine observable on-chain activity with elevated exposure to illicit sources.

However, these findings also partly can explain the results obtained in the network analysis of the previous section. Profiles 1 and 4 were most frequently represented in the network, which is consistent with the higher use of valid blockchain addresses. This increased on-chain activity may reflect the use of shared services, leading to more connections and the formation of clusters within the network. While such patterns can be indicative of potentially suspicious behavior, the results should be interpreted with caution, as the different analyses appear to reinforce rather than independently validate one another. Moreover, even though Profiles 2, 3, and 5 are less associated with blockchain transactions and show

lower levels of illicit exposure, no definitive conclusions can be drawn regarding their illicit potential. These profiles may potentially rely on alternative mechanisms, such as the voucher-based system Cryptex offered, which are not captured in the blockchain data. Furthermore, interestingly, higher percentages are observed for Profiles 2 and 3 compared to their presence in the withdrawal network depicted in Figure 8.1. Thus, it can be concluded that the profiles are not completely absent from the blockchain, but are less associated with address-sharing patterns.

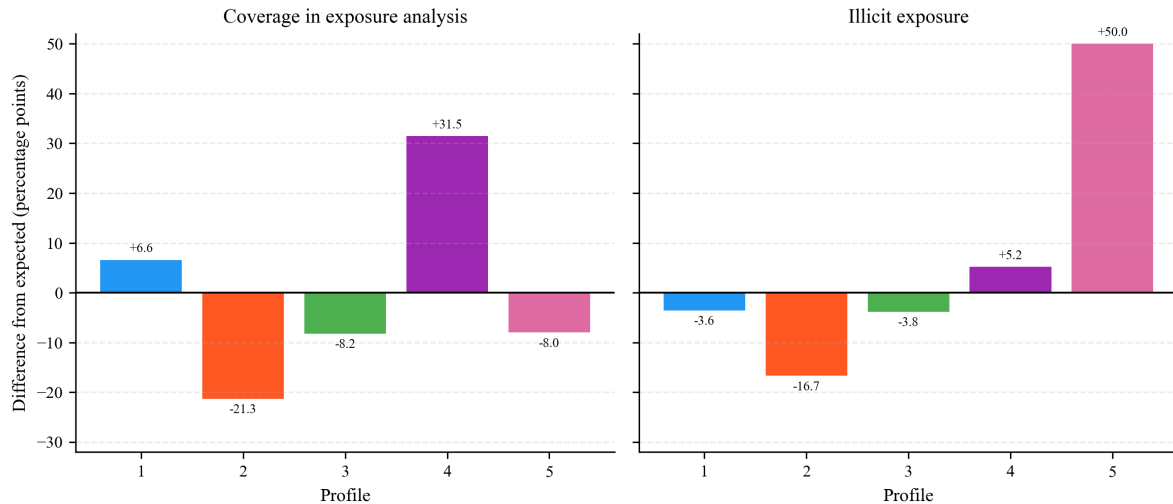


Figure 8.4: Over- and underrepresentation in blockchain coverage and illicit exposure across latent profiles

Figure 8.4 summarizes the results and shows that Profile 1 and 4 exhibit higher levels of blockchain coverage. Here, illicit exposure is accounted for coverage, meaning that based on blockchain presence, 50% of the users covered in the analysis show illicit exposure (see D.2). This large positive deviation observed for Profile 5 should be interpreted with caution, as the estimate is based on only two covered users, both of whom exhibited illicit exposure. Therefore, this observed deviation is not necessarily considered robust and is likely influenced by the small size of Profile 5.

8.4. Profiles and Attribution Mitigation

As a final part of this post-hoc analysis, identity attribution mitigation strategies (i.e., KYC-compliance) and financial attribution mitigation strategies (i.e., the use of privacy coin) are assessed. These are often part of specific transactional behavior (Financial Action Task Force, 2020). In doing so, it is assumed that these transactional behaviors patterns obtained through the LPA, may relate to the use of attribution mitigation mechanisms.

8.4.1. KYC-Compliance

This section looks at the distribution of KYC verification statuses across the profiles, using both the first and latest recorded statuses. Since these can be updated over time, both are included to provide a more complete view of user compliance. The results are presented in Figure 8.5 that shows a deviation plot on the latest KYC-verification statuses of users. Appendix section D.3 provides more detailed information on the distribution of statuses, including users their initial recorded status. For the interpretation in this chapter, focus is put on the latest KYC-statuses, as these are assumed to be the most representative. However, as indicated in section D.3, some differences exist when considering initial statuses, particularly for Profiles 2 and 3, which initially show higher proportions of 'Not approved' statuses (see section D.3).

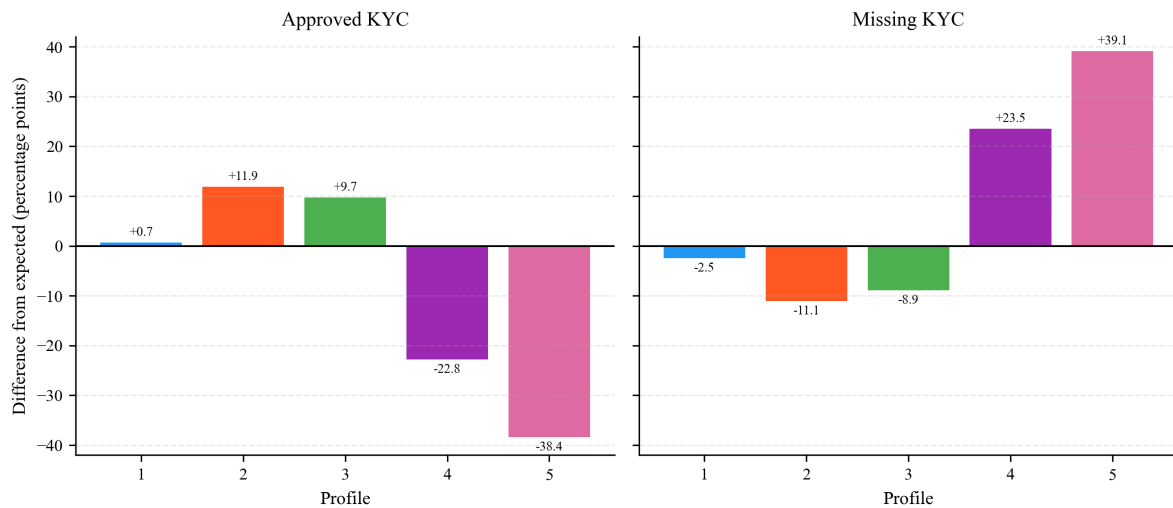


Figure 8.5: Over- and underrepresentation of latest KYC statuses across profiles

The distribution of KYC verification statuses in Table D.4 shows notable differences across the profiles. Profiles 1, 2, and 3 show relatively high proportions of approved users in both the first and latest KYC-status, suggesting a more compliant group of users, but this may also indicate that a potential greater willingness among these users to engage in KYC-procedures. In contrast, Profiles 4 and 5 have substantially higher shares of missing KYC-information, with more than half of the users lacking a final status. This may reflect a lower level of active engagement in compliance procedures, or the presence of internal users.

Profile 4 shows a relatively low proportion of approved users compared to other profiles, particularly in the initial KYC-status, although this improves slightly in the latest status. Profile 5 remains consistently low in approval rates, though interpretation is constrained by the small group size and high level of missing data.

All in all these patterns suggest that while some profiles are characterized by relatively complete and stable compliance information, others are marked by limited visibility, which may reflect differences in active behavior and the extent to which users intent to comply with regulations and engagement with the platform. However, an important sidenote is that it remains uncertain if Cryptex applied its KYC-procedures with integrity, or that approval outcomes may have been influenced by platform-specific biases.

8.4.2. Use of Privacy Coins

Figure 8.6 illustrates the deviation between the observed proportion of users within each profile that conducted at least one transaction involving the privacy coin Monero, an indicator often associated with illicit activity (AMLC, 2022), and the overall proportion observed in the dataset. This representation allows for a direct comparison with the expected distribution, and thereby highlighting profiles of which use of privacy coin Monero is either over- or underrepresented. Here, positive values indicate an over-representation relative to all users and negative values indicate that privacy coin use is underrepresented within a profile. The results show clear differences in the level of using a privacy-enhancing cryptocurrency across profiles.

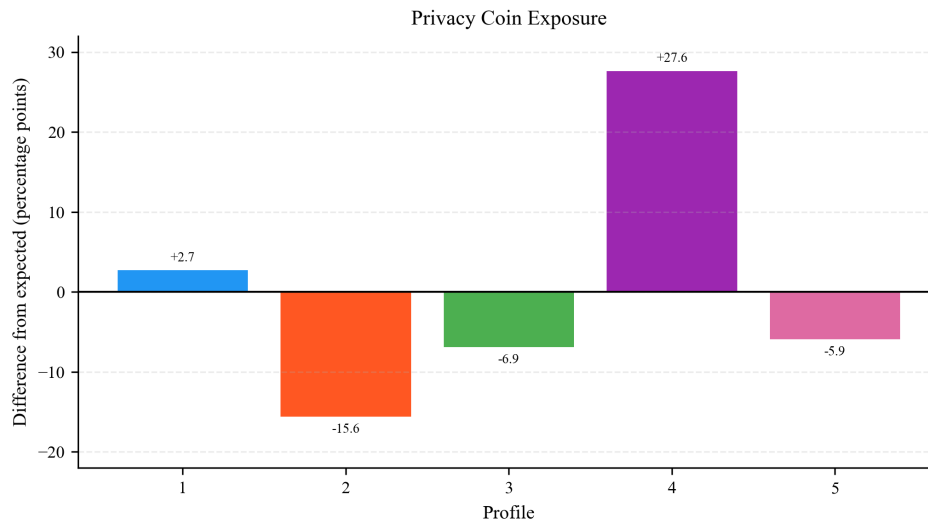


Figure 8.6: Over- and underrepresentation of privacy coins usage across profiles

Most notably, Profile 4 shows a strong overrepresentation, with a deviation of 27.6 percentage points compared to its share in the total study population. This suggests that users within this profile are substantially more likely to engage with privacy-enhancing cryptocurrency Monero. Profile 3 is moderately underrepresented (6.9 percentage points), while Profiles 1 and 5 are close to their expected shares. It should be noted however, that Profile 5 has a very small sample size ($N = 9$), which limits the interpretability of its deviation. In contrast, Profile 2 is relatively strongly underrepresented with -15.6 percentage points. This indicates that such behavior is relatively uncommon within this profile. Likewise, it is important to note that Profile 2 is characterized by low activity levels and only a single observation of privacy coin usage is observed in the total user group. Nevertheless, since the analysis considers presence instead of frequency, it may still suggest that less active accounts are likely to adopt privacy-enhancing cryptocurrencies within their single transactions.

Overall, the findings demonstrate that the use of privacy coin Monero is not evenly distributed across profiles, but rather is concentrated within specific behavioral patterns. In particular, Profile 4 suggests a clear alignment between characteristics captured by the LPA and behaviors associated with privacy-enhancing strategies.

8.5. Conclusion

This section combines the results to examine whether the identified profiles show clear differences across the conducted post-hoc analyses, as well as reflecting on the insights that can be gained, both descriptive and from an investigative perspective that contribute to the research objective of this thesis.

8.5.1. Consistency matrix

Below, Figure 8.7 displays the relative intensity of five indicators analyzed in this chapter for each identified profile. To do so, the results of the previous section have been translated by the researcher into an ordinal scale ranging from low (-2) to high (2). Therefore, colors within Figure 8.7 represent relative levels within the dataset, rather than absolute values.

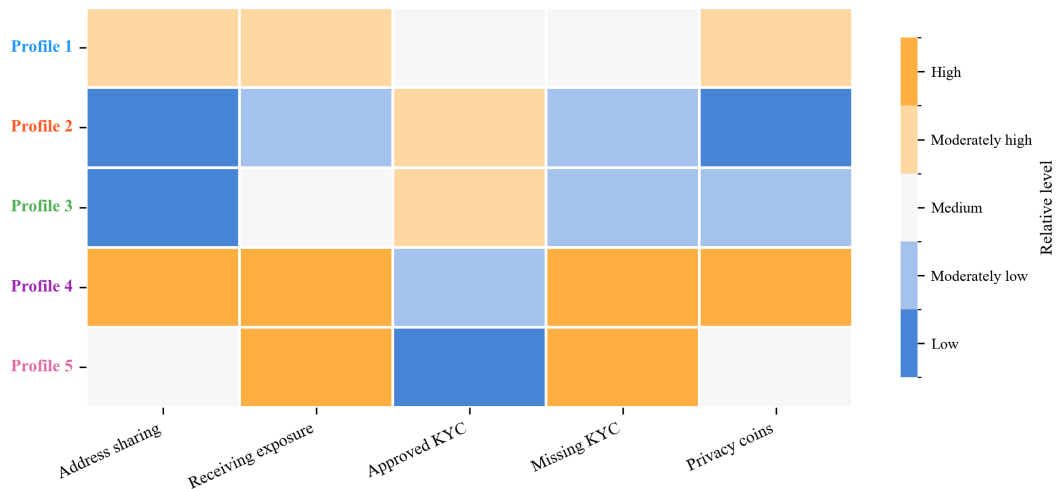


Figure 8.7: Comparative patterns of post-hoc analysis across profiles

- Profile 1:** Moderately Active Diversified Users
- Profile 2:** Extreme Passive Accounts
- Profile 3:** Low-Activity Users
- Profile 4:** High-Volume Active Traders
- Profile 5:** Systematic Accumulators

Based on Figure 8.7, the profiles reveal distinct patterns in terms of address-sharing behavior, exposure to illicit activity for receiving transactions within Cryptex, KYC-compliance, and use of a privacy coin. Together, these dimensions provide additional context to the obtained profiles by the LPA, suggesting that the profiles reflect meaningful differences in behavioral patterns.

Most notably in the analyses of this chapter is Profile 4, which is characterized by consistently high scores across multiple indicators, including address sharing, receiving exposure, missing KYC, and privacy coin usage. This coherent pattern suggests that this high-intensity behavioral cluster, may be indicative of structurally elevated risk due to frequent address-sharing, illicit exposure within their transactions and mitigation strategies for attribution. Profile 1 shows a similar pattern on almost all indicators of this post-hoc analysis, but with less intensity.

On the contrary, Profile 5 also less symmetric patterns which, however, can be associated with more elevated risk, combining high receiving exposure and missing KYC with low levels of approved KYC. This group appears highly active within Cryptex, but mainly transacts with a much less diversified portfolio of cryptocurrencies and relatively conducts a lot less withdrawals compared to Profile 4, potentially explaining their relative absence on the blockchain.

Profiles 2 and 3 show generally lower levels across most indicators, strong absence from address-sharing patterns regarding their withdrawals, with moderate variation in KYC-compliance. These profiles are less active overall, but still their overrepresentation in approved KYC-statuses remains striking when considering their relatively limited interaction with the platform. Profile 3 seems to transact more through the blockchain with Cryptex, also when comparing their presence on the blockchain, compared to the more active Profile 1.

These results illustrate how LPA on transactional behavioral indicators can be used to uncover meaningful variation in such datasets. At the same time, when interpreting the results, it should be emphasized that mentioned risk should not be directly interpreted as illicit intent, and observed differences may also reflect legitimate privacy preferences and thus similar patterns may still reflect different underlying intentions. Moreover, the observations and therefore difference are descriptive in nature. No formal statistical tests (e.g., chi-square tests for differences in proportions) have been conducted to assess the significance of the presented differences. Therefore the results should be interpreted as exploratory and indicative, and no definitive conclusions can be drawn regarding the strength or generalizability of

these patterns.

8.5.2. Impact of Internal Users

Within this thesis, often a distinction has been made in the analysis between regular and internal, administrative user account, to determine their presence and potential influence. A complete overview of results can be found in section C.2. Internal users are spread across all profiles, but are most prevalent within Profiles 1 and 4, which represent moderately and highly active user profiles. However, given the limited size of both the study population and the internal user sample, these comparisons should be interpreted as indicative rather than conclusive. However, as the numbers are very small, no sharp conclusions are drawn on their presence on potentially influencing the results and interpretation of this post-hoc analysis.

Qualitative Insights from Law Enforcement

This chapter presents the findings from the focus group conducted with law enforcement experts, which addresses SQ4: *“What insights can be derived from the identified profiles and do any of them display patterns consistent with criminal typologies discussed in the literature/interviews?”* and SQ5: *“How can the obtained results inform and support investigative strategies of law enforcement agencies in practice?”*

The focus group provides a substantive enrichment of the profiles and post-hoc contributions by discussing them with law enforcement. This complementary research phase aims to contextualize, expert validate, and critically reflect on the quantitative findings, as well as the applicability of LPA from an investigative perspective. The results are presented thematically, focusing on how experts interpret the profiles, assess their relevance for criminal typologies, and evaluate their applicability within operational practice.

The results were thematically analyzed based on a focus group summary, which can be found in Appendix E. The findings reveal four key themes that focus on the interpretation and practical relevance of the derived profiles, which highlight both the potential and limitations of the approach within an investigative context.

9.1. Challenges in Interpreting and Validating Latent Profiles

An important finding is that the current identified profiles are hard to interpret directly, and that this is highly dependent on contextual information. Across respondents, it became clear that transactional patterns alone are insufficient to draw conclusions about user intent or potential criminal involvement. It requires insight into the broader context in which these behaviors occur. This contextual information also requires to include the operational environment of the dataset, in particular the nature of the cryptocurrency exchange Cryptex as an important aspect. For example, the non-KYC nature of this exchange was repeatedly highlighted as an important factor, as it enables users to create multiple accounts easily. This significantly complicates the attribution of behavior and the interpretation of observed patterns, as this can also differ across exchanges and thus prior experience. Moreover, proper interpretation requires additional datapoints that are highly dependent on the criminal behavioral assumption that is being investigated by experts. All in all, the current derived profiles are deemed plausible, but impossible to directly interpret or label.

As a result, the profiles were not perceived as directly interpretable by the respondents, but rather as abstract representations that require additional context. The additional information is highly dependent on the operational setting of the dataset, as well as the criminal *modus operandi* under assumption. Respondents indicated that without this context, similar transactional patterns could reflect very different types of behavior or intentions. This seems to suggest that the value of the profiles does not lie in the

standalone interpretation of them and direct usage, but rather in how they can be combined with other information and investigative insight. This seems to be a fundamental shortcoming when applying LPA in an investigative context.

9.2. Dependence on Contextual Variables

Closely related to the challenges in interpreting and validating the profiles mentioned in the previous section, it can generally be concluded how important nuance and contextual information is in investigative research. Respondents emphasized a set of transactional dimensions that are usually and initially considered during investigative practices. These variables are commonly used for the initial screening and filtering of large datasets and were identified as essential for interpreting user behavior and are indeed represented in the current indicator selection of the LPA.

Most notably, transaction volume, the number of transactions, and the ratio between deposits and withdrawals were repeatedly mentioned as important behavioral dimensions. Moreover, temporal activity of users, such as bursts of activity within short periods of time, were highlighted as important signals of potentially suspicious behavior. The type of assets used, including the use of stablecoins or privacy coins, was also considered relevant, as it may provide insight into the intent or strategy of users. Perhaps it is therefore not the number of different assets used that is interesting, but the use of a specific asset. Furthermore, the conversion of funds into fiat currency was identified as a critical indicator, particularly in distinguishing whether funds remain within the crypto ecosystem or are integrated into the traditional financial system, a key element of money laundering processes.

Despite the fact that a large share of these dimensions are incorporated in the LPA of this research, still information is deemed missing that is also strongly dependent on the context of the case. Respondents indicated a strong need for additional data enrichment for this specific case of Cryptex. Specifically, interest was mentioned in variables such as the method of access (e.g., API versus application use), fiat on- and off-ramps, and potential links between multiple accounts. This highlights a key limitation of the current approach: while the profiles capture general behavioral patterns, they lack the detail required for direct meaningful interpretation and guidance for investigations.

9.3. Practical Relevance

The practical relevance of the profiles and usage of LPA as a tool, was found primarily in an exploratory, rather than operational use. One respondent from the specialized team clearly stated that the profiles would not be used as direct steering or decision-making tools in investigations, particularly not in contexts requiring evidentiary support or attribution. Instead, across all respondents the value was mainly seen in supporting an initial data exploration and forming initial ideas about what types of behavior or (criminal) patterns might be present in the dataset. For the AOT respondents, the profiles were perceived as a potentially useful tool or “rulebook” to guide the initial assessment of datasets. In this context, the profiles may assist in identifying whether a given account exhibits behavior that deviates from expected patterns and whether further investigation or escalation to specialized teams is considered necessary. As such, the profiles can function as a preliminary sorting tool, that can help structure large datasets into smaller, easier to handle subsets.

For more specialized investigators, the use case seemed to be more limited and nuanced. One respondent noted that the profiles could serve as a secondary validation tool: when investigating a specific account, the profile could be used to assess whether the observed behavior is in line with known patterns or appears more unusual. In this sense, the profiles contribute to contextualizing individual cases rather than directing investigative actions. Additionally, one specialized investigator mentioned that this type of information is always useful to learn from, although it may not be applied directly. Additionally, the profiles were considered useful for identifying similar behavioral patterns across accounts. Once a particular modus operandi has been identified, the profiles may help in locating other accounts exhibiting comparable behavior, thereby supporting pattern recognition within large datasets.

9.4. Implications for Investigative Research and Detection Bias

Another important insight from the focus group is that currently, investigative practices are strongly influenced by prioritization strategies which are based on expected impact. Respondents indicated that

investigations often focus on high-volume accounts, reflecting a tendency to “go after the big money.” This approach is driven by practical constraints and the expectation that such accounts are more likely to yield significant results.

However, one respondent indicated that this focus may introduce a certain bias in detection, potentially leading to not sufficiently investigating alternative behavioral patterns. In particular, low-volume or distributed transaction strategies, such as *smurfing*, may be overlooked due to their lower individual impact. Interestingly, profiles that are characterized by minimal or atypical activity (e.g., Profile 2) were also considered interesting for further investigation, contradicting the usual volume practices. Potentially, deducing behavioral patterns may aid law enforcement experts to reconsider their usual focus, suggesting that such profiles may capture behaviors or specific modus operandi that would otherwise remain undetected.

In this sense, the profiling approach may contribute to broadening the analytical perspective within investigations. By structuring the dataset into distinct behavioral patterns, the method encourages attention to a wider range of user behaviors, including those that do not conform to traditional high-volume indicators. This highlights the potential of the approach to complement existing investigative strategies by reducing bias in data selection and prioritization. In addition, by using more neutral behavioral indicators, the method may help uncover previously overlooked criminal modus operandi, particularly when the resulting profiles are enriched with contextual information that supports the assessment of illicit intent.

9.5. Critical Reflections and Limitations

As one of the final questions, respondents reflected critically on the limitations of LPA approach. A key issue, which can also be derived clearly from previous sections, is that the profiles are quite abstract. While they help to distinguish user groups, it results in a loss of nuance and detail. This makes the profiles difficult to interpret and provide value on their own, meaning that additional information is needed to really understand what they represent. Especially exposure to illicit transactions on the blockchain was considered missing, as well as specifically the type of virtual asset used.

Moreover, according to the respondents, it would be more valuable to filter on specific criminal modus operandi that are assumed to be present. For example, during the focus group, experts showed particular interest in the practice of *smurfing* or the potential use of multiple accounts, which do not uniformly apply to all types of exchanges.

One respondent also highlighted a fundamental trade-off between model complexity and interpretability. While the inclusion of additional variables could enhance the richness of the profiles, it may also reduce statistical robustness, particularly given the size and structure of the dataset. As such, the current profiles represent a balance between generalization and specificity, in which LPA may not fully align with the needs of investigative practice.

These limitations, combined with the subjective interpretation of the single researcher, should be taken into account when interpreting the findings from this qualitative research part and should be regarded as an initial exploration of law enforcement perspectives on the results and the potential of LPA in an investigative context.

9.6. Conclusion

Overall, the results show that, while the identified profiles are plausible, they are not easy to interpret, validate or directly useful on their own. Since the constructed profiles are quite abstract, they require additional context, such as the characteristics of the exchange and assumptions about possible criminal behavior, to make sense of them.

Instead, the main value of LPA lies in its exploratory use. It can potentially help to initially structure and filter large datasets and supports an initial understanding by highlighting patterns that may be interesting to investigate further. In this way, LPA as a method can mainly contribute to learning and forming initial ideas, rather than to directly identifying suspects or making decisions. Drawing conclusions about intent still depends on human expertise and additional data points, that are very hard to specify before conducting an LPA, as the LPA itself generates questions that are necessary to guide further analysis.

At the same time, using more neutral behavioral indicators, such as in this thesis, could allow for a broader view of the data. This can help find patterns that might otherwise be missed or under-prioritized when focusing only on high-volume accounts. However, to gain deeper insight into specific types of criminal behavior, more targeted and intent-driven variables may be needed. As a result, the value of LPA as a supporting tool strongly depends on the context and the type of user.

Finally, the usefulness of the profiles is strongly influenced by the context in which they are applied. In this case, factors such as the non-KYC nature of the exchange play an important role in shaping user behavior and its interpretation, which limits how well the findings can be generalized. Overall, LPA should be seen as a supporting tool for exploring and structuring data, rather than as a standalone method for operational decision-making. Human intelligence and contextual interpretation remain essential to translate the findings.

10

Discussion

This chapter reflects on the conducted research and its findings, in relation to the research objectives. This is followed by a discussion on the methodological considerations that impact this study and its limitations. Lastly, the implications for practice and further research are discussed.

10.1. Key Findings

10.1.1. Interpretation of the Profiles and Post-Hoc Analysis

In this thesis, a Latent Profile Analysis (LPA) was conducted to identify unobserved clusters of users that exhibit similar patterns in transactional behavior on the cryptocurrency exchange platform Cryptex. This provides a novel, data-driven approach within this specific research context, while retaining clarity on the underlying behavioral patterns. The resulting profiles from the analysis indicate that behavior within Cryptex is not homogeneous, but can be differentiated into distinct groups based on seven indicators that capture different behavioral dimensions to structure transactions. Differentiation was mainly driven by activity levels, varying use of assets and variation in transaction sizes. The obtained profiles not only show substantial variation in activity levels on the platform, but also notable differences in the complementary post-hoc analyses, which included address-sharing patterns on the blockchain for their withdrawals, exposure to illicit blockchain activity, and associations with identity mitigation strategies.

The presence of these clearly distinguishable profiles does indicate that LPA is effective in capturing variation in exchange-level transactional user behavior, while the post-hoc analyses suggest that these differences also relate to broader blockchain-related patterns and variables associated with increased risk. This underscores the notion in policy documents that risk-related behavior cannot be solely defined by a single indicator (Financial Action Task Force, 2020), but is a combination of interacting indicators.

Most notably the research shows the existence of a high-intensity profile (Profile 4), which is further characterized by a substantial higher share of blockchain related deposits and withdrawals, as well as elevated levels of address-sharing (indicating shared infrastructural usage patterns, common service usage, or potentially coordinated behavior), strategies associated with reducing traceability, and illicit exposure. Profile 5 also emerged as a notable profile during the post-hoc analysis and the focus group discussion. However, this profile is considered too unstable across the sensitivity analyses and due to its size, to support any robust or meaningful conclusions.

An important consideration here is that several of the post-hoc indicators are closely related to overall activity levels. Because users in Profile 4 substantially transact more and make more extensive use of blockchain-based services, they are also more likely to appear in exposure analyses and address-sharing networks. As a result, a part of the observed overrepresentation may reflect greater blockchain visibility rather than a higher underlying risk level.

On the contrary, profiles with generally lower levels across the LPA indicators point towards users that show limited engagement within Cryptex and more KYC-compliant behavior within the platform, but also seem to use different off-ramp strategies as they show limited behavior on the blockchain and are also not associated with patterns of address-sharing. However, we did not account for the lifespan

of a user its account. Although the burstiness indicator partially compensates for this, it may still result in biased interpretations, where users with shorter lifespans appear less active simply due to a more limited observation window.

Moreover, the post-hoc analysis mainly focuses on on-chain behavioral patterns. However, other cash-in and cash-out strategies, for example of lower-activity users, were not examined in further detail. As a result, the identified profiles may emphasize certain behavioral dimensions more strongly, while leaving other dimensions underrepresented and may influence the overall interpretation of the profiles and their relative interest. At the same time, the findings raise questions regarding the role of Cryptex as a non-KYC exchange platform. The absence of mandatory KYC-procedures may have contributed to higher transactional activity among certain users or attracted users seeking increased anonymity or reduced oversight. The specific profiles and their interpretation are therefore likely influenced by exchange-specific characteristics. However, this relationship cannot be directly established within the scope of the current analysis and therefore remains speculative.

Overall, these findings directly address the main research question and demonstrate that the identified profiles and their characterizations show meaningful differences, providing a structured lens to interpret user activity in such datasets. At the same time, the results should be interpreted by considering that certain behaviors associated with elevated risk may also reflect other, legitimate privacy considerations.

Although it was found that the profiles mainly split on activity levels, this limited "salsa effect" (Sinha et al., 2021) can still be valuable information for grouping users in purpose of investigation practice, as for most money laundering practices, activity is still considered a strong determinant for illicit behavior (Financial Action Task Force, 2020; Fu et al., 2025). It was found that activity levels do correlate with the account age of their users. This is not unsurprising, however, the current static representation of the profiles makes it difficult to capture potential transitions between profiles over time. As a result, the analysis does not account for how user behavior may evolve and can provide a biased image of a user, for instance from low-activity, exploratory usage toward more sustained or intensive activity patterns. This highlights a limitation of the current approach and suggests that longitudinal methods would be required to better capture such behavioral dynamics.

10.1.2. Integration with Findings from the Focus Group

The focus group results indicate that the profiles should not be interpreted as definitive classifications of user intent, but may arise due to different underlying motivations, which also depends on contextual factors. The findings emphasize that further characterization of the obtained profiles is necessary, as they remain relatively generic and therefore cannot be interpreted directly or standalone. However, from a law enforcement perspective, this enrichment strongly depends on the context of exchange use. This is in line with earlier statements from literature, that standalone analytical methods are generally considered insufficient in investigative contexts (Fu et al., 2025). Similarly, the findings from the focus group support the argument that LPA is most effective in combination with other analytical techniques and human expertise.

Furthermore, the applicability of the identified profiles depends on the role of the user within law enforcement and the nature of its investigations. The results suggest that LPA has several potential use cases: data filtering, limiting detective bias, attribution purposes or generalization. Of these use cases, the generalization of behavioral patterns aligns most closely with the methodological standards and strengths of LPA. This can largely be explained by the fact that LPA is primarily designed as a method to identify and generalize patterns within relatively homogeneous populations. This creates tension with investigative practice and applications, which typically focus on anomaly detection, but also requires contextual nuance and human intelligence which LPA cannot provide.

In this context, generalization could aid certain teams in developing structured, heuristic guidelines for identifying recurring behavioral patterns and associated profiles that may deserve additional attention. In that sense, LPA may support this by revealing recurring behavioral configurations, while remaining more adaptive to heterogeneous behavioral structures than rigid rule-based approaches. This shows similarities to a heuristic rule-based detection method mentioned by Fu et al. (2025), which are considered intuitive, easy to implement and cost-effective. However, unlike static rule-based approaches, the use of LPA allows for a more data-driven and behaviorally contextualized interpretation of the activity of users. Therefore, the identified profiles should not be translated into deterministic

classification rules without additional contextual validation and investigative interpretation.

This demonstrates how LPA can contribute to structuring complex transactional datasets and finding distinctive patterns, but to contribute meaning to these observations, a strong need for contextual interpretation in investigative settings remains necessary.

10.2. Methodological Reflection and Limitations

Data Limitations

First, there are several limitations regarding to the data used. The analysis relies on an empirically obtained dataset logged by Cryptex itself, which may introduce platform-specific biases and therefore not fully capture actual user behavior. From a research perspective, this limits control over the data generation process that can affect its quality. In addition, it should be taken into account that the analysis relies on the assumption that all recorded deposits and withdrawals included in the analysis correspond to actual transactions. Moreover, as this study focuses on deposit- and withdrawal-related behavior, it does not represent Cryptex its complete user population, nor its functionalities and thus the complete range of user activity.

LPA Tensions

Regarding methodological decisions, the construction of the selected indicator variables may influence the performance and interpretability of the model. While the chosen indicators were operationalized in such a way to capture relevant behavioral dimensions, their validity and robustness have not been formally evaluated. In particular, smaller or less representative samples may be more sensitive to how these indicators are defined, potentially affecting the stability of the resulting profiles.

In addition, the original indicators were deliberately retained, instead of applying a dimensionality reduction technique such as Principal Component Analysis (PCA), despite the presence of high correlations between variables. This could have potentially improved statistical efficiency and reduce the reliance on relaxing the local independence assumption, but would have reduced interpretability when only assessing the transactional dimension of behavior on the platform. In a law enforcement setting, it is important that profiles can be directly linked to observable transactional characteristics. Therefore, interpretability and practical relevance were prioritized over statistical parsimony.

In relation to the use of LPA in this research, another key methodological tension arises in how outliers are treated and how small clusters are allowed to be. For statistical approaches, outliers are often minimized to enhance model stability. However, in investigative contexts, these outliers may be of particular interest, as they indicate deviant, potentially illicit behavior. Therefore, in this research, they were retained throughout. This is also strongly related to the earlier described tension between the generalizing nature of LPA and the relevance of deviant behavior in investigative practice. In this research, the recommended minimum cluster size of 5% was not strictly adhered to (Sinha et al., 2021). It can be concluded that the applicability therefore strongly depends on the purpose of the analysis. If the objective is to identify behavioral patterns that can be generalized to a broader population, established LPA guidelines should be followed more closely and outliers may be removed. However, when the goal is exploratory data structuring or the identification of smaller sample-specific behavioral profiles, retention of outliers is preferred and deviation from these guidelines may be justifiable, but the mathematical implications of such deviations were not examined in this research.

In addition, it is important to note that LPA assumes that class membership explains the observed behavior. However, class membership is not only probabilistic in assigning a user to a specific group, but within this research also probabilistic in the sense that it is impossible to label profiles as inherently illicit, as similar behavioral patterns may arise from different underlying intentions. At most, they may indicate patterns associated with elevated risk, but this elevated risk can only be derived from supplementary, contextual information.

Additionally, as discussed when assessing the results in the previous section, the transactional behavioral indicators included in the analysis are inherently dynamic and are dependent on the timeframe of the dataset. As a result, changes in user behavior over time cannot be captured in the current approach, limiting the ability to detect the extent of and transitions between profiles or evolving behavioral strategies.

Post-Hoc & Network Analysis Considerations

For the post-hoc analysis, a hard classify-analyze approach was adopted. While this approach is justified based on the posterior membership probabilities, it is usually avoided because it does not account for classification uncertainty in the interpretation of the observed differences (Clark & Muthén, 2009). This limitation is especially for Profile 4, which has the second highest misclassification rate and therefore requires an even more cautious interpretation, while also being interpreted as a profile with elevated risk.

Moreover and logically, this post-hoc analysis is dependent on model specifications and the selected number of profiles. Other specifications yield alternative profile structures, which in turn influences the distribution of key indicators in this analysis, such as network characteristics, exposure and KYC-compliance. This highlights the sensitivity of the analysis to model specification and again, underscores the exploratory nature of the findings.

Specifically for the network analysis based on withdrawals, the constructed network only captures overlap in observable blockchain withdrawal behavior and does not represent direct fund flows or the complete transactional ecosystem of users. As a result, coordinated behavior that takes place through other cash-out methods, privacy-enhancing services, or off-chain interactions may remain invisible within the current network structure. The network is also static, meaning that changes in behavior and temporal overlap over time cannot be observed, requiring careful conclusions on the meaning of observed user overlap cannot be drawn. In addition, the current approach provides limited insight into the broader blockchain position of users outside the observable Cryptex environment. Local centrality metrics become a lot more valuable in such networks, as it indicates the positioning in actual transaction flows. Future research could therefore benefit from more extensive blockchain tracing and complementary the analysis of the transaction flows, as well as wallet clustering techniques. This can aid in obtaining a more complete understanding of user interactions and financial behavior within cryptocurrency ecosystems.

Focus Group Limitations

Concerning the qualitative part of this research, the focus group was conducted with a limited number of respondents, who were only briefly involved in and informed on the conducted research and their understanding of the thesis was not fully assessed. This may have affected how respondents interpreted and evaluated them, as well as how they were presented the information. Moreover, the respondents may not be fully representative for the broader perspective of other specialists and law enforcement agencies due to the sampling method. Additionally, the relatively short duration of the focus group limited the ability to assess all research details thoroughly.

10.3. Contributions

10.3.1. Methodological Contributions

The results illustrate an initial application of LPA as an interpretable profiling technique within a cryptocurrency investigative context, to support structuring and behavioral understanding. Although this study focuses on a single case study, the methodological approach may also be applicable to other Virtual Asset Service Providers (VASP) and can therefore contribute to the transparent identification of user segments within such environments.

Following the criteria established in Chapter 5, other large-scale transactional datasets which are characterized by behavioral heterogeneity, may be analyzed similarly on distinct behavioral patterns.

This extends the approaches adopted in earlier research and provides a more transparent and interpretable structuring approach of behavioral patterns, compared to earlier research primarily focused on machine-learning (ML) techniques in identifying and analyzing subpopulations.

Furthermore, this study contributes methodologically as it reflects on the tensions between LPA and investigative practice. These include trade-offs between interpretability and statistical efficiency, as well as the treatment of outliers, the relevance of smaller behavioral clusters, and the probabilistic nature of profile membership. By having discussed these tensions in the context of a real-world dataset, that contains noise and incomplete observability, this research contributes to a more grounded understanding of the practical application and limitations of LPA within investigative settings.

By combining LPA based on internal transactional data with additional indicators derived from other

dimensions, such as blockchain-related behavior and privacy-enhancing characteristics, this research contributes to an initial, multi-dimensional analytical approach for examining cryptocurrency exchange behavior. This study emphasizes the relevance and value of combining various datasources and types. Such an integrative perspective has been identified by Fu et al. (2025) as an important direction for future research, to which this thesis aims to contribute.

Moreover, the research design includes a mixed-method approach, which incorporates perspectives from enforcement professionals to the quantitative analyses. Whereas existing research within this domain often focuses primarily on technical or statistical analysis, the integration of practitioner perspectives in the interpretation and use of the obtained results remains relatively underexplored. This study therefore contributes by bridging analytical profiling with investigative interpretation and applicability.

10.3.2. Scientific Contributions

As an empirical contribution, this study provides insights based on a unique internal dataset of the seized cryptocurrency exchange Cryptex. Compared to earlier studies that primarily relied on public observable blockchain data and less interpretable ML-approaches, this thesis contributes to a deeper, more transparent understanding of how users interact within cryptocurrency exchange environments and the broader spectrum of their behavioral variation related to those platforms.

The findings demonstrate in a structural and statistically grounded manner that substantial behavioral heterogeneity exists in how cryptocurrency exchange users interact with the platform, which supports the notion that users cannot be treated as a homogeneous population within investigative analysis. The identified profiles differed considerably in transaction activity, while also showing differences in blockchain presence and shared infrastructure usage, as well as exposure to illicit activity and variables associated with attribution mitigation strategies. This suggests that potentially risk-related behavior is not uniformly distributed, but concentrates differently across user groups based on transactional behavior on such platforms.

Importantly, the results of this thesis demonstrate that observable, publicly observable blockchain data capture only a part of behavior and transactional strategies present within cryptocurrency exchange environments. Several profiles remained relatively less visible within blockchain-related indicators, while exhibiting distinct transactional behavior on the exchange platform itself. This indicates that blockchain-centered tracing and detection approaches may overlook relevant behavioral structures and transactional strategies that become visible when internal exchange data is analyzed. This highlights the relevance of combining data sources, as internal exchange data with blockchain analysis, since relying solely on either one in tracing may provide a partial representation of behavioral risk patterns.

Furthermore, the findings contribute to existing research by demonstrating that behavioral relevance within cryptocurrency exchange environments emerges through combinations of transactional, blockchain-related, and contextual characteristics, rather than isolated indicators alone. This contributes to a more nuanced and context-sensitive understanding of cryptocurrency exchange behavior and highlights the importance of interpreting potentially suspicious activity within its broader behavioral and operational context, also confirmed in the focus group.

10.3.3. Implications for Investigative Practice

The findings of this thesis may have several implications for law enforcement practice, specifically for intelligence-driven cryptocurrency investigations and the interpretation of large (exchange) datasets.

First and foremost, latent profiling may support investigators in the structuring of large operational datasets into more interpretable, behavioral user segments. This provides a clearer overview of common behavioral patterns present in a dataset. As a result, it may support early-stage filtering, investigative prioritization and decision-making in a transparent manner. In addition, the identified profiles may help investigators in identifying users with similarities in behavior, for example to compare with already identified or arrested individuals. This is particularly interesting as the findings suggest that suspicion within operational datasets is often initially based on a single indicator. Complementary to this, insights into the existing behavioral profiles may support the identification of alternative behavioral configurations, which enables investigators to test assumptions and hypotheses and countering investigative bias by shedding light on common behavioral configurations otherwise considered less likely.

Furthermore, the results indicate that some user profiles remain relatively invisible within blockchain-based indicators. However, this does not necessarily imply a complete absence of illicit risk. As cryptocurrency investigations often rely on blockchain tracing methodologies, this highlights the importance of combining multiple data sources, as certain behavioral patterns or risks otherwise may remain undetected. In that sense, the results suggest that LPA may not necessarily support the direct detection of illicit actors, but rather the identification of broader behavioral structures and transactional configurations that may require additional investigative attention.

However, as particularly emphasized in the qualitative part of this research, contextual interpretation and human expertise remain highly essential. Human intel is considered important for capturing nuances, exceptions and ethical considerations that intelligence-based policing may fail to account for. As a result, the use of such profiles therefore becomes more limited for concrete attribution purposes and applicability of these use cases is dependent on the stage and objective of the investigation. Translating these probabilistic patterns into rigid rule-based heuristics introduces a substantial risk of false positives, as similar transactional behaviors may reflect fundamentally different underlying intents. Moreover, the static nature of such heuristics struggles to adapt to evolving money laundering techniques, which are inherently dynamic and context-dependent. Nevertheless, the results of this thesis demonstrate that behavioral profiling approaches have potential as a supportive analytical tool within intelligence-driven cryptocurrency investigations.

10.4. Recommendations for Future Research

Further research could extend this study in several directions. Here, the most evident starting points are listed. First, expanding the applied methodology across multiple exchanges could improve generalizability of the and provide a more comprehensive understanding of user behavior on these platforms and common patterns of interaction, as well as the interaction with exchange-specific characteristics since this thesis also argued that user behavior is strongly dependent on the context of the exchange (e.g. being a non-KYC exchange). This suggests that generalization may only be possible if the chosen indicator variables for the LPA are sufficiently robust and independent of platform-specific characteristics.

Second, future research could explore the application of LPA on a subset of users rather than the full dataset. In this current study, the analysis may be dominated by low-intensity or exploratory accounts still, the results mainly show the most variation in the magnitude of the indicators. Focusing on a specific subset may reveal more fine-grained behavioral patterns within these segments.

Third, as mentioned in this discussion, incorporating longitudinal approaches, such as group trajectory modeling, would provide insights into whether users adopt evolving behavioral strategies and transition between different profiles over time, which is strongly assumed as it has been found that behavior is dependent on how long a user has been active on the platform.

Another promising direction is the integration of broader blockchain network analysis. Extending the scope beyond exchange-level data could provide deeper insight into user positioning within the blockchain and potentially within a money laundering network, which was now approximated solely by address-sharing patterns and network exposure.

Additionally, future research could focus on validation using ground truth data. While this study relied on expert validation by means of a focus group, incorporating known cases would allow for a more direct assessment of the relationship between identified profiles and actual illicit activity, which now remains very speculative within this research.

Finally, future research could integrate law enforcement practice further into the research design given this identified gap in current literature. Future studies could explore the investigative workflows in further detail, as well as corresponding needs of law enforcement for a more grounded assessment of the practical utility of LPA, in combination with other analytical tools.

Overall, this study is concluded to be very exploratory in nature. Future work could build on its findings by refining the research design for conducting an improved LPA, by improving indicator selection, and applying more advanced modeling approaches to enhance both robustness and practical relevance.

11

Conclusion

This thesis examined whether distinct behavioral profiles can be identified among the users of a seized cryptocurrency exchange and how these profiles can be of value within an investigative context, as formulated in the following main research question:

“To what extent can latent behavioral profiles be identified within the internal records of a seized cryptocurrency exchange, and what insights do these profiles provide for law enforcement investigations?”

The findings demonstrate that through the application of a Latent Profile Analysis (LPA) using several dimensions of transactional activity, users are not homogeneous in their behaviors but can be meaningfully differentiated into distinct profiles. These profiles mainly diverge in terms of activity intensity, variability in transaction sizes and the extent to which different cryptocurrencies were exchanged.

To evaluate their discriminative capacity and investigative relevance from a money laundering perspective, a post-hoc analysis was conducted to examine whether the identified groups also differ in other relevant dimensions related to financial crime risk. By doing so, particular attention was given to proxies related to network positioning and exposure to illicit activity, as well as KYC-compliance and the use of privacy coins.

The results indicate that the identified profiles extend beyond mere differences in transactional intensity, and are also associated with distinct behavioral patterns. Notably, one high-intensity profile consistently exhibited stronger blockchain integration, lower KYC-compliance, and greater exposure to risk-related indicators. More broadly, higher levels of transactional activity on the Cryptex platform are associated with increased on-chain activity and a greater overlap with external wallet addresses, indicating reliance on shared blockchain services or withdrawal infrastructures. In contrast, another high-activity profile, concentrated in a single asset, showed a substantially different network structure, while exhibiting complete direct receiving exposure to illicit activity.

Conversely, lower activity profiles demonstrate different external behavior strategies. Despite their limited activity on the exchange itself, these users do not demonstrate comparable levels of on-chain interaction, which potentially indicates alternative cash-in or cash-out mechanisms that are not directly captured within the scope of the post-hoc analysis. This suggests that observable blockchain activity captures only part of the transactional strategies present within the exchange environment. In addition, they appear more likely to comply with KYC-requirements.

However, the identified profiles are inherently platform-specific, as they are derived from the internal dynamics and user base of a single seized exchange. This limits the extent to which the findings can be directly generalized to other platforms or contexts, particularly given the influence of exchange-specific characteristics such as the absence of KYC-requirements.

A focus group conducted with experts from law enforcement further supports the exploratory value of this research and provides insight into the potential value of this type of statistical modeling in practice. Respondents indicated that the profiles are not directly interpretable in isolation, but instead provide

a useful abstraction to help structure large volumes of transactional data and increase awareness of behavioral patterns, potentially limiting detection bias. However, human intelligence remains essential for law enforcement investigations, as it can incorporate nuances or details that statistical modeling cannot. As such, the profiles were perceived most valuable depending on the phase and objective of an investigation and when used in combination with additional information and expertise, rather than as a standalone indicator of illicit activity.

In answering the main research question, this study demonstrates that latent profiles derived from internal exchange data can reveal distinct transactional behavior patterns that are indicative of different user strategies, including those potentially associated with elevated financial crime risks. While the profiles themselves do not directly identify illicit activity, they offer a structured way to reduce complexity within large transactional datasets and can support the direction and learning potential of investigative efforts.

In doing so, this study provides a first exploratory foundation for applying LPA to systematically uncover meaningful behavioral structures within cryptocurrency exchange datasets, contributing to more context-driven, behaviorally informed, and data-driven approaches to financial crime investigations in cryptocurrency environments.

Statement on the use of AI

This page includes a statement on the use of AI by the author.

While preparing this work, several AI-based tools have been used. Specifically, ChatGPT and Claude have been adopted as brainstorming partners and to receive critical feedback on content ideas, as well as the grammar of sections where I considered improvement was needed. This is considered to have improved the overall structure and argumentation of this thesis, including my own thought processes as the author. Moreover, ChatGPT and Claude have been used to support in writing code for analyses in SQL and Python, while the actual dataset remained separate and content was not shared with these tools. Additionally, I used Grammarly to prevent spelling errors and improve my English proficiency, while maintaining my original writing tone. After using these tools and services, I thoroughly reviewed and edited the content as necessary. Accordingly, I take full responsibility for the content of this thesis document.

References

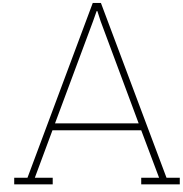
- Aigner, A., & Dhaliwal, G. (2021). UNISWAP: Impermanent Loss and Risk Profile of a Liquidity Provider. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3872531>
- Aitken, C., Roberts, P., & Jackson, G. (2010). Fundamentals of probability and statistical evidence in criminal proceedings. 36.
- Alexander, C., Chen, X., Deng, J., & Fu, Q. (2023). Market Efficiency Improvements from Technical Developments of Decentralized Crypto Exchanges. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4495589>
- AMLC. (2022, January). Signalen van witwassen via cryptocurrency. Retrieved December 17, 2025, from <https://www.amlc.nl/indicatoren-voor-het-witwassen-van-geld-via-cryptocurrency/>
- Anti-Money Laundering Network. (n.d.). Cryptocurrency Laundering Database. Retrieved December 8, 2025, from <https://amlnetwork.org/watchdog-database/cryptocurrency-laundering/>
- Antoniuk, D. (2024). Russia arrests nearly 100 with suspected ties to sanctioned crypto exchange. <https://therecord.media/russia-arrests-cybercriminals-cryptex-crypto>
- Antonopoulos, A. M. (2017). *Mastering Bitcoin: Programming the open blockchain* (Second edition) [OCLC: 990142087]. O'Reilly.
- Bank for International Settlements. (2022). BIS Annual Economic Report 2022. <https://www.bis.org/publ/arpdf/ar2022e.htm>
- Belastingdienst. (2024, September). Seizure of 7 million euros of crypto currency and 2 crypto currency exchanges offline. Retrieved September 11, 2025, from <https://www.fiod.nl/seizure-of-7-million-euros-of-crypto-currency-and-2-crypto-currency-exchanges-offline/>
- Biscontin, T. (2024). Cryptocurrency exchange (digital currency exchange). <https://www.ebsco.com/research-starters/business-and-management/cryptocurrency-exchange-digital-currency-exchange>
- Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, 29(2), 213–38. <https://doi.org/10.1257/jep.29.2.213>
- Booij, T. M., Verburgh, T., Falconieri, F., & Wegberg, R. S. V. (2021). Get Rich or Keep Tryin' Trajectories in dark net market vendor careers. *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 202–212. <https://doi.org/10.1109/EuroSPW54576.2021.00028>
- Bryant, R., & Kennedy, I. (2014, January). Investigating digital crime.
- Business Wire. (2021, September). Cryptex – an Anonymous Cryptocurrency Exchange Has Updated Its Blog Structure. <https://www.businesswire.com/news/home/20210920005681/en/Cryptex-an-Anonymous-Cryptocurrency-Exchange-Has-Updated-Its-Blog-Structure>
- Candeloro, L., Savini, L., & Conte, A. (2016). A New Weighted Degree Centrality Measure: The Application in an Animal Disease Epidemic (R. Lambiotte, Ed.). *PLOS ONE*, 11(11), e0165781. <https://doi.org/10.1371/journal.pone.0165781>
- Chainalysis. (2019a, August). Chainalysis launches first actionable Real-Time alerts for suspicious transactions across 15 cryptocurrencies. <https://www.chainalysis.com/blog/real-time-alerts-press-release/>
- Chainalysis. (2019b, October). The 5 Key Types of Cryptocurrency Exchanges. Retrieved December 8, 2025, from <https://www.chainalysis.com/blog/cryptocurrency-exchange-types/>
- Chainalysis. (2020). *Who's Who On The Blockchain* (tech. rep.). https://go.chainalysis.com/rs/503-FAP-074/images/Typologies-Report-final.pdf?mkt_tok=NTAzLUZBUC0wNzQAAAGenGrqtxeRIHrqjkdJEI43n2Q7KC1Y_D8-88naW2dP6FbS1kt_nwJ4VUVYZMj8kiXokaV9HwnF6R8Cz0J6nCqizEmh43QciftANPrwnFI
- Chainalysis. (2022, April). Is Bitcoin traceable? <https://www.chainalysis.com/blog/is-bitcoin-traceable/>
- Chainalysis. (2023, November). Indirect Exposure: Why You Need to Look Beyond Direct Counterparties to Understand Cryptocurrency Address Risk. <https://www.chainalysis.com/blog/cryptocurrency-risk-blockchain-analysis-indirect-exposure/>

- Chainalysis. (2024, September). OFAC Designates Russian Exchange Cryptex, FinCEN names PM2BTC. <https://www.chainalysis.com/blog/ofac-sanctions-russian-exchange-cryptex-uaps-fraud-shop-2024/>
- Chainalysis. (2025, January). 2025 Crypto Crime Trends from Chainalysis. Retrieved September 17, 2025, from <https://www.chainalysis.com/blog/2025-crypto-crime-report-introduction/>
- Chin, C.-L., & Yao, G. (2014). Convergent Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1275–1276). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_573
- Clark, S., & Muthén, B. (2009). Relating Latent Class Analysis Results to Variables not Included in the Analysis.
- Conlon, T., Corbet, S., & Hu, Y. (2023). The collapse of the FTX exchange: The end of cryptocurrency's age of innocence. *The British Accounting Review*, 101277. <https://doi.org/10.1016/j.bar.2023.101277>
- Conti, M., Sandeep Kumar, E., Lal, C., & Ruj, S. (2018). A Survey on Security and Privacy Issues of Bitcoin. *IEEE Communications Surveys & Tutorials*, 20(4), 3416–3452. <https://doi.org/10.1109/COMST.2018.2842460>
- Creswell, J., & Clark, V. (2011). *Designing and Conducting Mixed Methods Research*. SAGE Publications. <https://books.google.nl/books?id=YcdIPWPJRBcC>
- Crypto.com. (2024, September). What Is an Order Book and How Does It Work? Retrieved December 17, 2025, from <https://crypto.com/nl/university/what-is-a-crypto-order-book>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dutch Financial Intelligence Unit. (n.d.). The Dutch Approach to Prosecuting Money Laundering. Retrieved September 30, 2025, from <https://www.amlc.eu/the-dutch-approach-to-prosecuting-money-laundering/>
- Dutch Ministry of Justice and Safety. (2024, June). Summary Cybercrime Monitor Netherlands [Last Modified: 2024-06-10T16:25]. Retrieved September 17, 2025, from <https://www.om.nl/documenten/publicaties/cybercrime/2024/ccbn/samenvatting-ccbn-2024-engels>
- Europol. (2022). *Cryptocurrencies: Tracing the evolution of criminal finances* (Europol Spotlight Series). <https://www.europol.europa.eu/publications-events/publications/cryptocurrencies-tracing-evolution-of-criminal-finances#downloads>
- Europol. (2024, April). Cryptocurrencies and financial crime: A strategic approach to ensure security. Retrieved September 17, 2025, from <https://www.europol.europa.eu/media-press/newsroom/news/cryptocurrencies-and-financial-crime-strategic-approach-to-ensure-security>
- Fanusie, Y., & Robinson, T. (2018). Bitcoin laundering: An analysis of illicit flows into digital currency services. *Center on sanctions and illicit finance memorandum, January*.
- Financial Action Task Force. (2020). *Virtual Assets Red Flag Indicators of Money Laundering and Terrorist Financing* (tech. rep.). Paris. <https://www.fatf-gafi.org/en/publications/Methodsandtrends/Virtual-assets-red-flag-indicators.html>
- Fu, Q., Liu, J. K., Pan, S., & Yuen, T. H. (2025). SoK: A Deep Dive Into Anti-money Laundering Techniques for Blockchain Cryptocurrencies. In W. Susilo & J. Pieprzyk (Eds.), *Information Security and Privacy* (pp. 310–330). Springer Nature Singapore.
- Galletta, A., & Cross, W. E. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication* (Vol. 18). NYU press.
- Goh, K.-I., & Barabási, A.-L. (2008). Burstiness and memory in complex systems. *Europhysics Letters*, 81(4), 48002. <https://doi.org/10.1209/0295-5075/81/48002>
- Golbeck, J. (2013, January). Chapter 3 - Network Structure and Measures. In J. Golbeck (Ed.), *Analyzing the Social Web* (pp. 25–44). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-405531-5.00003-1>
- Gu, Z., Lin, D., & Wu, J. (2022). On-chain analysis-based detection of abnormal transaction amount on cryptocurrency exchanges. *Physica A: Statistical Mechanics and its Applications*, 604. <https://doi.org/10.1016/j.physa.2022.127799>

- Guest, G., & Fleming, P. (2015, January). Mixed Methods Research. <https://doi.org/10.4135/9781483398839.n19>
- Hägele, S. (2024). Centralized exchanges vs. decentralized exchanges in cryptocurrency markets: A systematic literature review. *Electronic Markets*, 34(1), 33. <https://doi.org/10.1007/s12525-024-00714-2>
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural equation modeling: a multidisciplinary journal*, 14(2), 202–226.
- Hu, Y., Sun, Y., Wu, L., Zhou, Y., & Chang, R. (2024). Towards Understanding and Analyzing Instant Cryptocurrency Exchanges. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 8(3). <https://doi.org/10.1145/3700430>
- Hu, Y., Wang, W., & Liu, Y. (2025). Deep Learning Empowered Blockchain Transaction Prediction and Anomaly Detection. 2277 *CCIS*, 50–61. https://doi.org/10.1007/978-981-97-9412-6_5
- J5. (2024). Crypto Assets Risk Indicators. <https://www.irs.gov/pub/irs-ci/j5-media-release-5-24-2024.pdf>
- Jiang, H., Zhang, K., Ma, X., Sun, Y., & Ma, Y. (2022). Suspicious Customer Detection on the Blockchain Network for Cryptocurrency Exchanges. 1679 *CCIS*, 265–277. https://doi.org/10.1007/978-981-19-8043-5_19
- Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), 602. <https://doi.org/10.2307/2392366>
- Johnson, Onwuegbuzie, & Turner, L. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1, 112–133. <https://doi.org/10.1177/1558689806298224>
- Kim, W.-W., Yang, S.-O., Lee, J., Suh, B., Kim, J., Kang, S.-Y., & Hong, K. (2025). AI-Based Framework for Classifying Cryptocurrency Exchanges through Transaction Feature Analysis. <https://doi.org/10.1109/ICBC64466.2025.11114507>
- Kruisbergen, E., Leukfeldt, E., Kleemans, E., & Roks, R. (2019). Money talks money laundering choices of organized crime offenders in a digital age. *Journal of Crime and Justice*, 42, 1–13. <https://doi.org/10.1080/0735648X.2019.1692420>
- Lee, J., Kim, W. -., Suh, B. W., Shin, Y., & Kim, J. (2025). Unveiling the Behavioral Pattern of Cryptocurrency Exchange Services based on Transaction Analysis [Journal Abbreviation: 2025 7th International Conference on Blockchain Computing and Applications (BCCA)]. *2025 7th International Conference on Blockchain Computing and Applications (BCCA)*, 68–75. <https://doi.org/10.1109/BCCA66705.2025.11229498>
- Leung, F.-H., & Savithiri, R. (2009). Spotlight on focus groups. *Canadian Family Physician Medecin De Famille Canadien*, 55(2), 218–219.
- Leuprecht, C., Jenkins, C., & Hamilton, R. (2022). Virtual money laundering: Policy implications of the proliferation in the illicit use of cryptocurrency. *Journal of Financial Crime*, 30(4), 1036–1054. <https://doi.org/10.1108/JFC-07-2022-0161>
- Levi, M. (2015). Money for Crime and Money from Crime: Financing Crime and Laundering Crime Proceeds. *European Journal on Criminal Policy and Research*, 21, 275–297.
- Levi, M., & Soudijn, M. (2020). Understanding the Laundering of Organized Crime Money. *Crime and Justice*, 49, 000–000. <https://doi.org/10.1086/708047>
- Lubbertsen, K., van Eeten, M., & van Wegberg, R. (2025). Ghost Clusters: Evaluating Attribution of Illicit Services through Cryptocurrency Tracing, 1357–1374.
- Magidson, J., & Vermunt, J. K. (2005). A Nontechnical Introduction to Latent Class Models.
- Miedema, F., Lubbertsen, K., Schrama, V., & Wegberg, R. v. (2023). Mixed Signals: Analyzing Ground-Truth Data on the Users and Economics of a Bitcoin Mixing Service. *32nd USENIX Security Symposium (USENIX Security 23)*, 751–768. <https://www.usenix.org/conference/usenixsecurity23/presentation/miedema>
- Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 76–86.
- Morrison, D., Bedinger, M., Beevers, L., & McClymont, K. (2022). Exploring the raison d’être behind metric selection in network analysis: A systematic review. *Applied Network Science*, 7(1), 50. <https://doi.org/10.1007/s41109-022-00476-w>
- Morse, J. M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Routledge.

- Moser, M., Bohme, R., & Breuker, D. (2013, September). *An inquiry into money laundering tools in the Bitcoin ecosystem* [Pages: 14]. <https://doi.org/10.1109/eCRS.2013.6805780>
- Muthén, B., & Muthén, L. K. (2000). Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling With Latent Trajectory Classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882–891. <https://doi.org/10.1111/j.1530-0277.2000.tb02070.x>
- Nakamoto, S. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. *Cryptography Mailing list at https://metzdowd.com*.
- Nylund-Gibson, K., Asparouhov, T., & Muthén, B. (2007). Deciding On the Number of Classes In Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Copyright*, 14, 535–569. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., & Choi, A. (2018). Ten Frequently Asked Questions About Latent Class Analysis. *Translational Issues in Psychological Science*, 4, 440–461. <https://doi.org/10.1037/tps0000176>
- Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from Latent Classes: A Demonstration of Different Approaches to Include Distal Outcomes in Mixture Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6), 967–985. <https://doi.org/10.1080/10705511.2019.1590146>
- Ostern, N. K., & Riedel, J. (2021). Know-Your-Customer (KYC) Requirements for Initial Coin Offerings. *Business & Information Systems Engineering*, 63(5), 551–567. <https://doi.org/10.1007/s12599-020-00677-6>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, n160. <https://doi.org/10.1136/bmj.n160>
- Pettersson Ruiz, E., & Angelis, J. (2022). Combating money laundering with machine learning – applicability of supervised-learning algorithms at cryptocurrency exchanges. *Journal of Money Laundering Control*, 25(4), 766–778. <https://doi.org/10.1108/JMLC-09-2021-0106>
- PR Newswire. (2021). Cryptex App - anonymous gate to the world of crypto can now exchange USD to bitcoin or vice versa. <https://www.prnewswire.co.uk/news-releases/cryptex-app-anonymous-gate-to-the-world-of-crypto-can-now-exchange-usd-to-bitcoin-or-vice-versa--846627785.html>
- Rabade, R., Mishra, N., & Sharma, S. (2014). Survey of Influential User Identification Techniques in Online Social Networks [Series Title: Advances in Intelligent Systems and Computing]. In S. M. Thampi, A. Abraham, S. K. Pal, & J. M. C. Rodriguez (Eds.), *Recent Advances in Intelligent Informatics* (pp. 359–370, Vol. 235). Springer International Publishing. https://doi.org/10.1007/978-3-319-01778-5_37
- Reynolds, P., & Irwin, A. (2017). Tracking digital footprints: Anonymity within the bitcoin system: *Journal of Money Laundering Control*, 20(2), 172–189. <https://doi.org/10.1108/JMLC-07-2016-0027>
- Riccardi, M., & Levi, M. (2018, May). Cash, Crime and Anti-Money Laundering. https://doi.org/10.1007/978-3-319-64498-1_7
- Schrama, V., Laarschot, J. v., Volten, C., & Wegberg, R. v. (2022). Virtuele valuta. *Handelingsperspectieven voor data-gedreven opsporing*. <http://hdl.handle.net/20.500.12832/3215>
- Scott, J. (2000). *Social Network Analysis: A Handbook* 2nd Ed.
- SEBA Bank. (2020, June). *A Beginner's Guide to Blockchain Accounting Standards* (tech. rep.). Retrieved December 9, 2025, from <https://aminagroup.com/wp-content/uploads/2020/06/a-beginners-guide-to-blockchain-accounting-standards--seba.pdf>
- Shcherbak, S. (2014). How Should Bitcoin Be Regulated? https://ejls.eui.eu/wp-content/uploads/sites/32/pdfs/Spring_Summer2014/HOW_SHOULD_BITCOIN_BE_REGULATED-%20Sergii_Shcherbak_.pdf
- Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. *Critical Care Medicine*, 49(1), e63–e79. <https://doi.org/10.1097/CCM.0000000000004710>
- Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of Vocational Behavior*, 120, 103445. <https://doi.org/10.1016/j.jvb.2020.103445>

- St. Vincent and the Grenadines Financial Services Authority. (2025, May). Implementation of the Virtual Asset Business Act & Commencement of the Registration Process for Virtual Asset Businesses. Retrieved December 8, 2025, from <https://fsasvg.com/implementation-of-the-virtual-asset-business-act-commencement-of-the-registration-process-for-virtual-asset-businesses/>
- Toufaily, E., & Zalan, T. (2024). In blockchain we trust? Demystifying the “trust” mechanism in blockchain ecosystems. *Technological Forecasting and Social Change*, 206, 123574. <https://doi.org/10.1016/j.techfore.2024.123574>
- Tovanich, N., Coquidé, C., & Cazabet, R. (2025). Cryptocurrency Network Analysis [Version Number: 1]. <https://doi.org/10.48550/ARXIV.2502.03411>
- TRM Labs. (2024, September). US Treasury Takes Coordinated Actions Against Illicit Russian Virtual Currency Exchanges and Cybercrime Facilitators PM2BTC and Cryptex. <https://www.trmlabs.com/resources/blog/us-treasury-takes-coordinated-actions-against-illicit-russian-virtual-currency-exchanges-and-cybercrime-facilitators-pmbtc-and-cryptex>
- United States Attorney’s Office. (2024, September). Two Russian nationals charged in connection with operating billion-dollar money laundering services; Justice Department seizes web domains for multiple illicit crypto exchanges. <https://www.justice.gov/usao-edva/pr/two-russian-nationals-charged-connection-operating-billion-dollar-money-laundering>
- van Wegberg, R., Oerlemans, J.-J., & van Deventer, O. (2018). Bitcoin money laundering: Mixed results?: An explorative study on money laundering of cybercrime proceeds using bitcoin. *Journal of Financial Crime*, 25(2), 419–435. <https://doi.org/10.1108/JFC-11-2016-0067>
- van de Laarschot, J., & van Wegberg, R. (2021). Risky Business? Investigating the Security Practices of Vendors on an Online Anonymous Market using Ground-Truth Data, 4079–4095. <https://www.usenix.org/system/files/sec21-van-de-laarschot.pdf>
- Van Lissa, C. J., Garnier-Villarreal, M., & Anadria, D. (2024). Recommended Practices in Latent Class Analysis Using the Open-Source R-Package tidySEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 31(3), 526–534. <https://doi.org/10.1080/10705511.2023.2250920>
- Van Wegberg, R., Miedema, F., Akyazi, U., Noroozian, A., Klievink, B., & Van Eeten, M. (2020). Go See a Specialist? Predicting Cybercrime Sales on Online Anonymous Markets from Vendor and Product Characteristics. *Proceedings of The Web Conference 2020*, 816–826. <https://doi.org/10.1145/3366423.3380162>
- Vermunt, J. K. (2004). Toepassingen van latent klasse analyse in sociaal wetenschappelijk onderzoek. *Sociale Wetenschappen*, 47, 2–14. <https://www.jeroenvermunt.nl/sw2004.pdf>
- Vermunt, J., & Magidson, J. (2002). Latent Class Cluster Analysis. In *Applied latent class analysis*. Cambridge University Press. <https://research.tilburguniversity.edu/en/publications/latent-class-cluster-analysis/>
- Volten, C., van Eeten, M., & van Wegberg, R. (2025). Money for Nothing, Supervision for a Fee: Investigating the Effects of the 5th Anti-Money Laundering Directive on Cryptocurrency Exchanges in the Netherlands. *European Journal on Criminal Policy and Research*. <https://doi.org/10.1007/s10610-025-09640-1>
- Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 46(4), 287–311. <https://doi.org/10.1177/0095798420930932>
- Whelan, C., Martin, J., Dupont, B., & Harkin, D. (2025). ‘The name of the game’: Policing perspectives on cybercrime disruption. *Police Practice and Research*, 1–21. <https://doi.org/10.1080/15614263.2025.2523522>
- Williams, G. A., & Kibowski, F. (2016). Latent class analysis and latent profile analysis.
- Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in psychology*, 5, 920.
- Zhou. (2025). Regulating Crypto Money Laundering: An Assessment of Current Regulatory Responses and Potentials for Technology-Based Solutions. <https://stanford-jblp.pubpub.org/pub/crypto-laundering/release/1>
- Zhou, F., Chen, Y., Zhu, C., Jiang, L., Liao, X., Zhong, Z., Chen, X., Chen, Y., & Zhao, Y. (2024). Visual Analysis of Money Laundering in Cryptocurrency Exchange. *IEEE Transactions on Computational Social Systems*, 11(1), 731–745. <https://doi.org/10.1109/TCSS.2022.3231687>
- Zola, F., Medina, J., & Orduna-Urrutia, R. (2025). Assessing the Impact of Sanctions in the Crypto Ecosystem: Effective Measures or Ineffective Deterrents? 15263 *LNCS*, 292–308. https://doi.org/10.1007/978-3-031-82349-7_20



Literature review search strategy

To identify the academic knowledge gap of Chapter 1, an initial literature review has been conducted by employing the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020 Page et al. (2021). This ensures transparency and reproducibility of the review process.

The primary database used for the conducted review is *Scopus*. Additionally, the same search query has been used within the ACM Digital Library, but this yielded no relevant literature. A complete flowchart of the process is depicted in Figure A.1 on the next page. The results from Scopus were retrieved on October 1st, 2025.

The initial search resulted in 27 publications after removing duplicates and excluding document types other than journal articles and conference proceedings. After an initial screening and assessment of eligibility, 10 publications remained from the Scopus database and were included in the literature review. Papers that primarily addressed legal or regulatory frameworks were excluded, as this study focuses on the behavioral aspects of financial crime instead of governance or compliance perspectives.

In addition, 1 source was found through snowballing and deemed relevant, and 5 sources were added based on input from affiliated researchers in the cybercrime domain or snowballing, which were mainly related to the use of Latent Class Models in the cybercrime domain, or provided a systematic body of knowledge. A separate search for (cybercrime OR "financial crime" OR "money laundering") AND ("latent class" OR "latent profile") resulted in one of those sources, and the others were considered irrelevant, as they mainly focused on victim profiling.

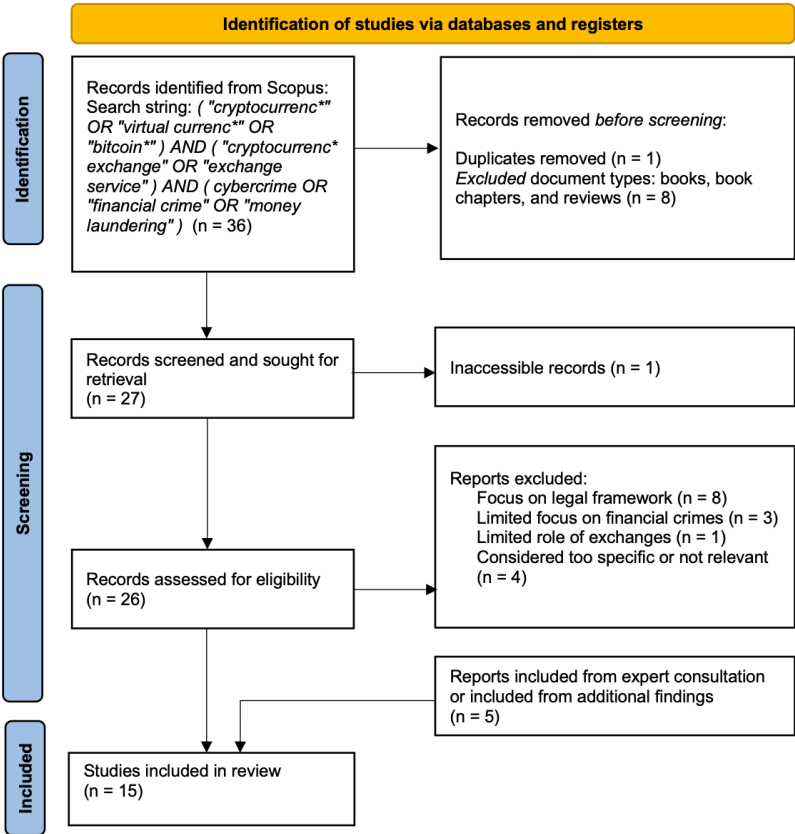


Figure A.1: Literature review flowchart PRISMA (2020)

A.1. Literature synthesis

The academic relevance and research question result from knowledge gaps identified through the 15 selected papers that follow from A.1 and are presented in Table A.1.

The focus is on previous research addressing financial crime through cryptocurrency exchanges. Table A.2 provides an overview of the analytical dimensions covered in the selected studies. It highlights that most research has mainly focused on automatic detection on open-source data, with little attention for behavioral and practical aspects. These observations form the basis of three knowledge gaps of section 1.2 that support the formulation of the main research question in subsection 1.3.3.

No.	Authors	Year	Title
1	Kim et al.	2025	<i>AI-Based Framework for Classifying Cryptocurrency Exchanges through Transaction Feature Analysis</i>
2	Van Wegberg et al.	2018	<i>Bitcoin money laundering: mixed results? An explorative study on money laundering of cyber-crime proceeds using bitcoin</i>
3	Pettersson Ruiz & Angelis	2022	<i>Combating money laundering with machine learning – applicability of supervised-learning algorithms at cryptocurrency exchanges</i>
4	Reynolds & Irwin	2017	<i>Tracking digital footprints: anonymity within the bitcoin system</i>
5	Hu et al.	2025	<i>Deep Learning Empowered Blockchain Transaction Prediction and Anomaly Detection</i>
6	Booij et al.	2021	<i>Get Rich or Keep Tryin': Trajectories in dark net market vendor careers</i>
7	Van Wegberg et al.	2020	<i>Go See a Specialist? Predicting Cybercrime Sales on Online Anonymous Markets from Vendor and Product Characteristics</i>
8	Gu et al.	2022	<i>On-chain analysis-based detection of abnormal transaction amounts on cryptocurrency exchanges</i>
9	Van de Laarschot & Van Wegberg	2021	<i>Risky Business? Investigating the Security Practices of Vendors on an Online Anonymous Market using Ground-Truth Data</i>
10	Jiang et al.	2022	<i>Suspicious Customer Detection on the Blockchain Network for Cryptocurrency Exchanges</i>
11	Hu et al.	2024	<i>Towards Understanding and Analyzing Instant Cryptocurrency Exchanges</i>
12	Zhou et al.	2024	<i>Visual Analysis of Money Laundering in Cryptocurrency Exchanges</i>
13	Zola et al.	2025	<i>Assessing the Impact of Sanctions in the Crypto Ecosystem: Effective Measures or Ineffective Deterrents?</i>
14	Lee et al.	2025	<i>Unveiling the Behavioral Pattern of Cryptocurrency Exchange Services based on Transaction Analysis</i>
15	Fu et. al	2025	<i>SoK: A Deep Dive Into Anti-money Laundering Techniques for Blockchain Cryptocurrencies</i>

Table A.1: Overview of included publications

No.	Authors	Year	ML	Open source	Behavioral aspects	Inclusion of practical field
1	Kim et al.	2025	✓	✓		
2	Van Wegberg et al.	2018		✓	✓	
3	Pettersson Ruiz & Angelis	2022	✓			
4	Reynolds & Irwin	2017		✓		✓
5	Hu et al.	2025	✓	✓		
6	Booij et al.	2021		✓	✓	
7	Van Wegberg et al.	2020		✓	✓	
8	Gu et al.	2022	✓	✓		
9	Van de Laarschot & Van Wegberg	2021			✓	
10	Jiang et al.	2022	✓	✓		
11	Hu et al.	2024		✓		
12	Zhou et al.	2024	✓			✓
13	Zola et al.	2025		✓	✓	
14	Lee et al.	2025		✓	✓	
15	Fu et al. *	2025	✓	✓	✓	✓

Table A.2: Overview of included literature and analytical dimensions.

Note. * Paper 15 is a summarizing paper, addressing AML-detection methods and current research practices. Therefore it discusses, yet it does not apply ML or uses open source data.

B

LPA: Detailed Results

B.1. Correlations of Indicator Variables

To assess the suitability of the indicator variables for the latent profile analysis, pairwise Spearman correlations were computed and examined, as presented in Figure B.1. The circles displayed in Figure B.1 indicate that the bivariate correlation exceeds ($\rho > 0.7$), which indicates potential redundancy between indicators. In particular, indicators related to transaction activity (e.g., number of deposits, withdrawals, and active days) and transaction size (e.g., average deposit and withdrawal values) exhibited substantial overlap. From both a mathematical and behavioral perspective, these correlations are considered expected and interpretable. Although there is no universally defined threshold for indicator removal identified in literature and LPA models also capitalize on these correlations, in this thesis ($\rho > 0.7$) are assessed for removal on the condition that all behavioral dimensions classified in Chapter 5 are still represented in the conceptual model of the LPA. This is considered important, as remaining these indicators may give too much weight to a specific activity dimension, in which the model will be forced to capitalize, which is expected not provide as much behavioral nuance.

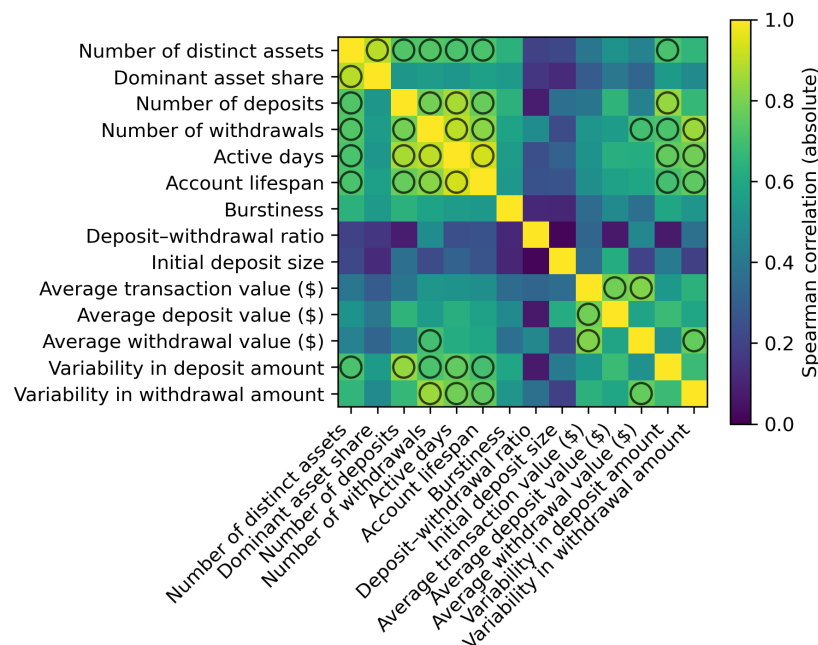


Figure B.1: Correlations between potential indicator variables
Note. Circle overlays represent a correlation between variables higher than 0.7.

The final selection of indicators, of which the correlation matrix is presented in Figure B.2, includes variables representing transaction intensity (number of deposits), temporal dynamics (burstiness), relative behavior (deposit–withdrawal ratio), account characteristics (initial deposit size), asset concentration (dominant asset share), and transaction magnitude and variability.

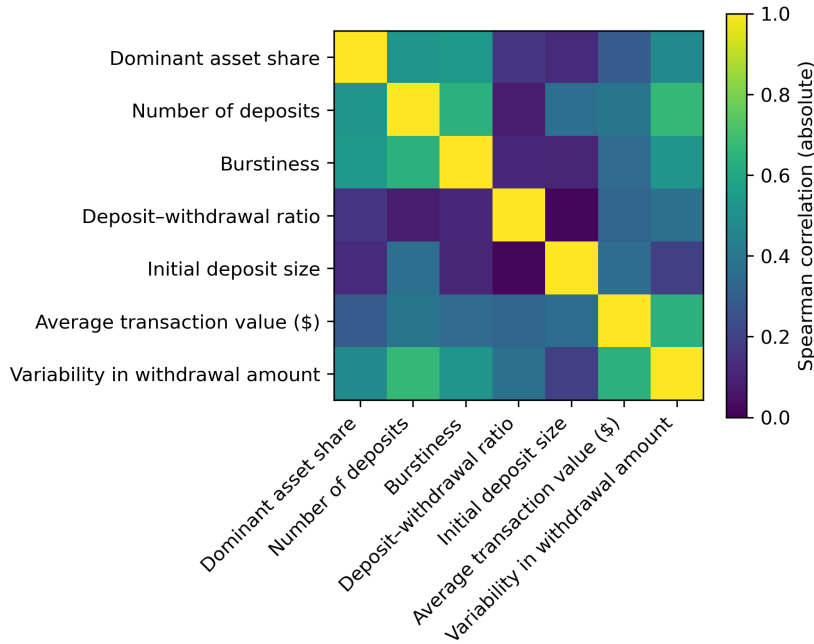


Figure B.2: Correlations between the final selection of indicator variables

The resulting correlations in Figure B.2 show no excessively high correlations, which suggests that the selected indicators provide a better balance, while capturing the key dimensions of user behavior relevant for the clustering analysis.

B.2. Relaxation of the local independence assumption

An important aspect of LPA is the assumption of local independence between indicators within each latent class (Sinha et al., 2021). Inspection of the bivariate residuals (BVRs) of the results showed that none of the estimated models resulted in all BVR values falling below the conventional chi-square threshold of 3.84.¹ This indicates that within a class, indicators still share information that can not be explained by the latent class structure. However, relatively large BVRs were observed, which may indicate substantial residual dependence between indicators and potentially affect class formation. Highest BVR-values were found between average transaction value (USD) and the standard deviation of withdrawals, as well as between the number of deposits and the deposit-to-withdrawal ratio. These dependencies are conceptually plausible and therefore do not appear unexpectedly, as transaction volume and variability are inherently related, and the deposit-to-withdrawal ratio is mathematically linked to the transaction counts. To assess the influence of this, residual covariances were allowed between these respective pairs. Below, the results of this are presented in Table B.1.

¹The value 3.84 corresponds to the critical value of the chi-square distribution with one degree of freedom at $\alpha = .05$.

Table B.1: Model fit indices for LPA solutions - without relaxation of the local independence assumption

Model	LL	BIC	AIC	Npar	Max. BVR	Entropy R^2
1-Cluster	-2562.6030	5202.9474	5153.2060	14	116.1276	1.0000
2-Cluster	-1932.5606	4026.1570	3923.1211	29	41.5439	0.9937
3-Cluster	-1723.0538	3690.4379	3534.1077	44	14.9909	0.9506
4-Cluster	-1544.1867	3415.9980	3206.3734	59	20.9248	0.9543
5-Cluster	-1428.2944	3267.5079	3004.5889	74	16.5412	0.9648
6-Cluster	-1334.5830	3163.3795	2847.1661	89	11.9977	0.9777
7-Cluster	-1259.6835	3096.8749	2727.3671	104	11.5331	0.9754
8-Cluster	-1197.5091	3055.8204	2633.0182	119	9.5169	0.9740
9-Cluster	-1131.1054	3006.3075	2530.2109	134	7.6401	0.9748
10-Cluster	-1054.7456	2936.8822	2407.4913	149	8.1816	0.9784

Note. LL = log-likelihood; BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion.

B.3. Log-Likelihood Stability Diagnostics

Table B.2: Log-likelihood values across random start configurations

Profiles	250 starts	500 starts	1000 starts
1	-2480.0751	-2480.0751	-2480.0751
2	-1844.3828	-1844.3828	-1844.3828
3	-1583.6988	-1583.6988	-1583.6988
4	-1409.7745	-1401.3948	-1401.3948
5	-1294.2654	-1309.3374	-1299.6624
6	-1212.6816	-1194.8406	-1195.9623
7	-1155.4904	-1131.7109	-1123.5356
8	-1086.4895	-1064.7159	-1066.9061
9	-1017.1039	-1015.0188	-1032.4912
10	-967.3117	-948.4212	-957.9619

Note. All models are estimated with 500 iterations.

Values represent the best log-likelihood obtained across random start configurations.

Table B.2 shows that the one- to three-profile models produce identical log-likelihood values across multiple random start configurations, indicating that these are stable solutions for which the global maximum has most likely been identified. For the 4-class model, the model requires more random starts to find the highest model fit. From five profiles onwards, the solutions become less consistent across random starts, indicating a higher likelihood of convergence to local maxima. As a result, it cannot be certainly concluded that these solutions represent a true global optimum within the study population.

B.4. Profile Solutions for the 3-, 4- and 5-Cluster Models

Tables Table B.3, Table B.4 and Table B.5 show the cluster means for each indicator variable and provide insight into the differences and further breakdown of clusters between respectively the 3-, 4- and 5-profile solution.

Table B.3: Standardized indicator means for the 3-profile solution

Indicator	C1	C2	C3
Cluster size	0.50	0.27	0.23
Transaction Frequency	0.59	−0.78	−0.41
Average Transaction Volume (USD)	0.50	−0.55	−0.48
Transaction Variability	0.88	−0.91	−0.89
Deposit–Withdrawal Ratio	−0.17	0.08	0.28
Burstiness	0.43	−0.60	−0.26
Dominant Asset Share	−0.45	0.96	−0.13
Initial Deposit	0.18	−0.32	−0.02

Note. Values represent standardized indicator means (z-scores) within each latent profile.

Table B.4: Standardized indicator means for the 4-profile solution

Indicator	C1	C2	C3	C4
Cluster size	0.41	0.26	0.22	0.11
Transaction Frequency	0.68	−0.78	−0.43	0.16
Average Transaction Volume (USD)	0.55	−0.54	−0.53	0.25
Transaction Amount Variability	0.90	−0.91	−0.90	0.52
Deposit–Withdrawal Ratio	−0.31	0.08	0.30	0.38
Burstiness	0.67	−0.59	−0.22	−0.66
Dominant Asset Share	−0.76	0.96	−0.17	0.90
Initial Deposit	0.22	−0.32	−0.08	0.10

Note. Values represent standardized indicator means (z-scores) within each latent profile.

Table B.5: Standardized indicator means for the 5-profile solution

Indicator	C1	C2	C3	C4	C5
Cluster size	0.28	0.26	0.23	0.19	0.04
Transaction Frequency	0.18	−0.79	−0.41	1.18	0.58
Average Transaction Volume (USD)	0.08	−0.54	−0.47	1.05	0.61
Transaction Amount Variability	0.63	−0.91	−0.89	1.17	1.09
Deposit–Withdrawal Ratio	−0.16	0.08	0.28	−0.52	1.53
Burstiness	0.09	−0.59	−0.28	1.04	−0.20
Dominant Asset Share	−0.38	0.96	−0.08	−0.81	0.88
Initial Deposit	−0.10	−0.32	−0.01	0.77	−0.90

Note. Values represent standardized indicator means (z-scores) within each latent profile.

Based on the results in the tables presented, it can be noticed that increasing the number of clusters from three to five leads to more detailed differentiation of profiles, rather than emergence of completely new patterns.

Throughout all solutions, the second cluster is remained, indicating a high distinctive, and stability towards this profile. Mainly cluster 1 in the 3- and 4-profile solution is prone to be further subdivided.

The 5-profile solution introduces a fifth, relatively small cluster characterized by a very high deposit-to-withdrawal ratio, implying much more deposits by these accounts than withdrawals. Moreover, the 4th cluster in Table B.5 can be interpreted as a more pronounced subgroup of C1 in Table B.4 with a remarkably high initial deposit and transaction frequency and burstiness.

Overall, the results of these tables suggest a hierarchical structure in which additional clusters capture increasing within-group heterogeneity, with transaction frequency, variability, and asset concentration as the key behavioral dimensions for distinction, which remain consistent across all three model specifications.

B.5. Distribution of Removed Indicator Variables

To avoid multicollinearity and overrepresentation of certain transactional dimensions in the LPA, a subset of indicator was excluded as previously discussed in section B.1. Figure B.3 presents the mean z-score distributions of these indicators across the identified profiles.

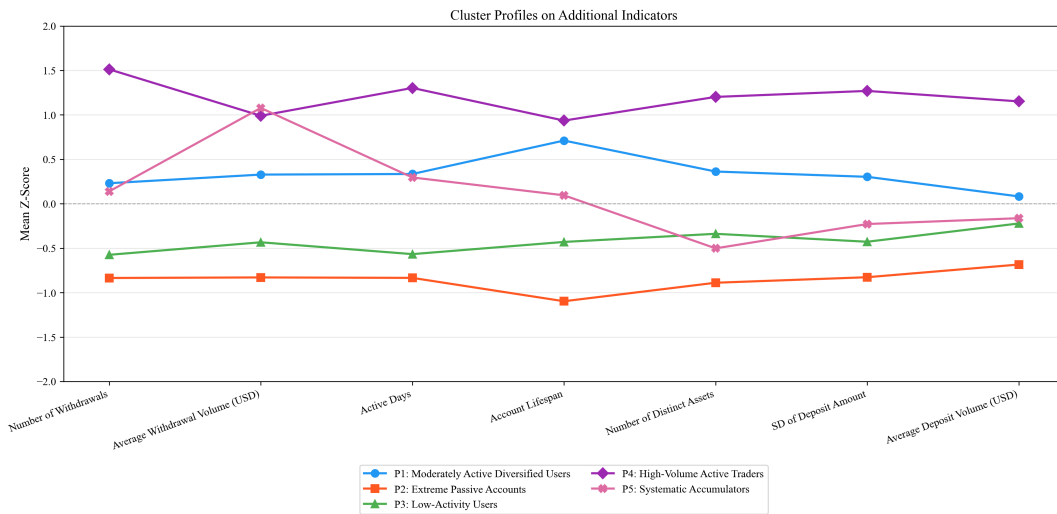


Figure B.3: Mean z-scores for removed indicator variables across the 5-profile solution

Figure B.3 shows a more pronounced "salsa-effect" (Sinha et al., 2021), in which the mean z-score profiles show largely parallel patterns across the, in this case, removed indicators. This suggests that these removed indicators do not provide substantial additional discriminatory power beyond those included in the analysis. Moreover, the plot suggests that the higher activity profiles (1, 4 and 5) are associated with an overall longer account lifespan. This is important to consider, as it suggests that there is a chance that users may transition between profiles over time, for example moving from initial low-intensity or exploratory behavior toward more active usage patterns, a behavioral pattern that is also reflected in existing literature (Toufaily & Zalan, 2024).

B.6. Sensitivity Analysis

This appendix section provides more detailed results on the conducted sensitivity analyses, respectively by removing outliers that are outside the 1.5 IQR range and by removing accounts that are presumed to be part of Cryptex its internal, administrative processes.

B.6.1. Outlier Removal

Table B.6: Model fit statistics — outlier removal sensitivity analysis (500 random starts)

Model	LL	BIC	AIC	Max. BVR	Entropy R^2
Model 1	-2271.04	4629.43	4574.08	89.21	1.0000
Model 2	-1680.85	3541.87	3427.71	22.54	0.9918
Model 3	-1431.57	3136.13	2963.15	10.82	0.9846
Model 4	-1324.93	3015.64	2783.85	7.89	0.9793
Model 5	-1241.78	2942.16	2651.55	7.18	0.9838
Model 6	-1174.23	2899.88	2550.46	5.05	0.9607
Model 7	-1127.18	2898.59	2490.36	4.68	0.9610
Model 8	-1052.74	2842.52	2375.48	3.15	0.9705
Model 9	-995.51	2820.87	2295.01	3.49	0.9795
Model 10	-946.82	2816.30	2231.63	2.41	0.9741

Note. Outliers were removed prior to estimation; all models were estimated using 500 random starts and 500 iterations.

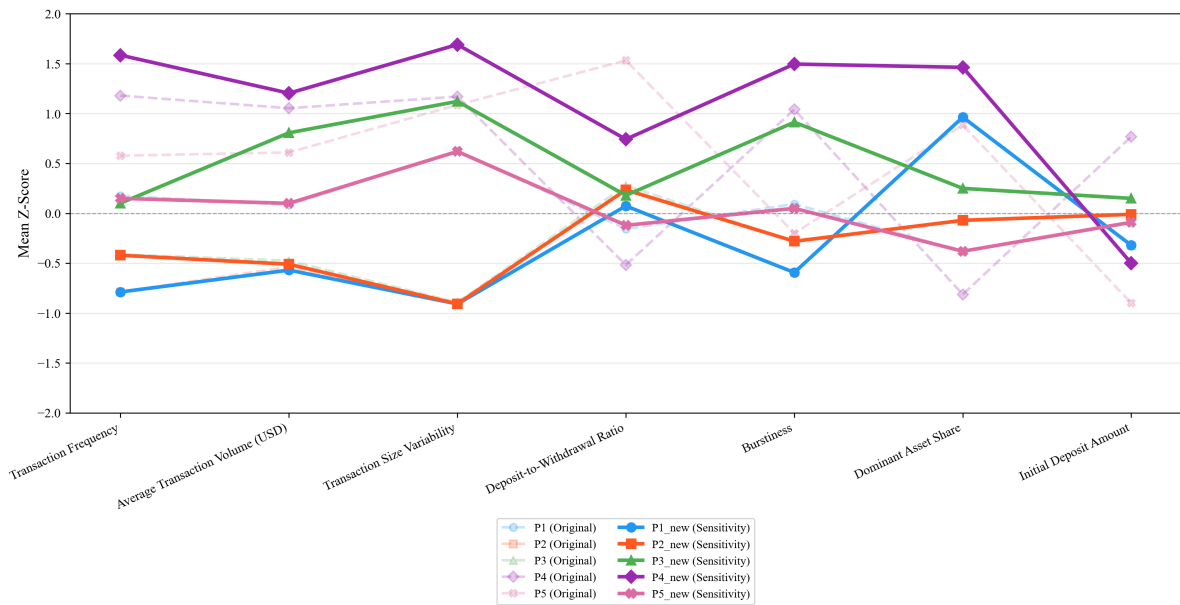


Figure B.4: Comparison of standardized indicator means between the original and outlier sensitivity analysis of the 5-profile solutions

Table B.7: Standardized indicator means for the 5-profile solution — outlier removal sensitivity analysis (500 random starts)

Indicator	C1	C2	C3	C4	C5
Cluster size	0.3697	0.2794	0.2133	0.0987	0.0390
Transaction Frequency	0.9321	−0.8436	−0.4365	0.3330	−1.2761
Average Transaction Volume (USD)	0.5666	−0.4818	−0.5144	0.1567	0.5031
Transaction Amount Variability	0.9784	−0.8779	−0.8611	0.3407	0.8731
Deposit–Withdrawal Ratio	−0.2565	0.3123	0.3796	−0.1082	−1.6383
Burstiness	0.8996	−0.6841	−0.2104	−0.6339	−0.9048
Dominant Asset Share	−0.8652	0.9492	−0.1563	0.8597	0.0986
Initial Deposit	0.2021	−0.2990	−0.0424	0.6564	−1.2174

Note. Values represent standardized indicator means (z-scores) within each latent profile.

Here, Table B.7 and Table B.5 are compared to draw conclusions about the robustness of the original latent profile analysis. This is visualized in Figure B.4.

The comparison between the original and outlier removal solutions of Figure B.4 shows that the overall behavioral structure remains largely similar after excluding outliers. The comparison indicates that several behavioral profile patterns remain observable after removing outliers. Although the overall profile shapes are broadly comparable, some differences in indicator intensity and profile separation can be observed. Most differences are observed for Profiles 4 and 5, with respect to the dominant asset share and initial deposit amount. This visualization shows that the removal of outliers has an influence on the distribution and extremity of certain behavioral characteristics, but the overall latent behavioral structure remains largely similar. This supports the overall robustness of the initial model including all observations.

B.6.2. Internal User Account Removal

Table B.8: Model fit statistics — internal user account removal sensitivity analysis (500 random starts)

Model	LL	BIC	AIC	Max. BVR	Entropy R ²
Model 1	−2285.50	4658.83	4603.01	93.41	1.0000
Model 2	−1674.79	3530.71	3415.57	26.75	0.9914
Model 3	−1430.87	3136.19	2961.74	13.97	0.9823
Model 4	−1267.39	2902.54	2668.78	8.88	0.9605
Model 5	−1165.68	2792.42	2499.35	6.98	0.9433
Model 6	−1072.84	2700.06	2347.68	6.24	0.9658
Model 7	−1005.16	2658.02	2246.33	5.13	0.9562
Model 8	−953.91	2648.83	2177.82	4.87	0.9641
Model 9	−891.70	2617.73	2087.41	3.41	0.9655
Model 10	−827.85	2583.32	1993.69	4.12	0.9716

Note. Internal users were removed prior to estimation; all models were estimated using 500 random starts and 500 iterations.

Table B.9: Standardized indicator means for the 5-profile solution — internal user account removal sensitivity analysis (500 random starts)

Indicator	C1	C2	C3	C4	C5
Cluster size	0.2645	0.2456	0.2377	0.2270	0.0252
Transaction Frequency	−0.7884	−0.1051	1.0020	−0.4235	1.5853
Transaction Amount Variability	−0.9081	0.4126	1.1224	−0.9075	1.6905
Deposit–Withdrawal Ratio	0.0686	−0.1990	−0.4519	0.2350	2.7425
Initial Deposit	−0.3498	−0.0473	0.5899	−0.0264	−0.7961
Burstiness	−0.5934	−0.2554	0.9133	−0.3358	1.4924
Dominant Asset Share	0.9637	−0.0444	−0.8768	−0.0043	0.4634
Average Transaction Volume (USD)	−0.5687	0.0562	0.8072	−0.5094	1.2037

Note. Values represent standardized indicator means (z-scores) within each latent profile.

Table B.9 is compared to Table B.5 draw conclusions about the behavioral structure being remained. Because direct comparison of both tables is difficult to inspect visually, the mean z-scores per profile are plotted against those of the original 5-profile solution in Chapter 7 below in ??.

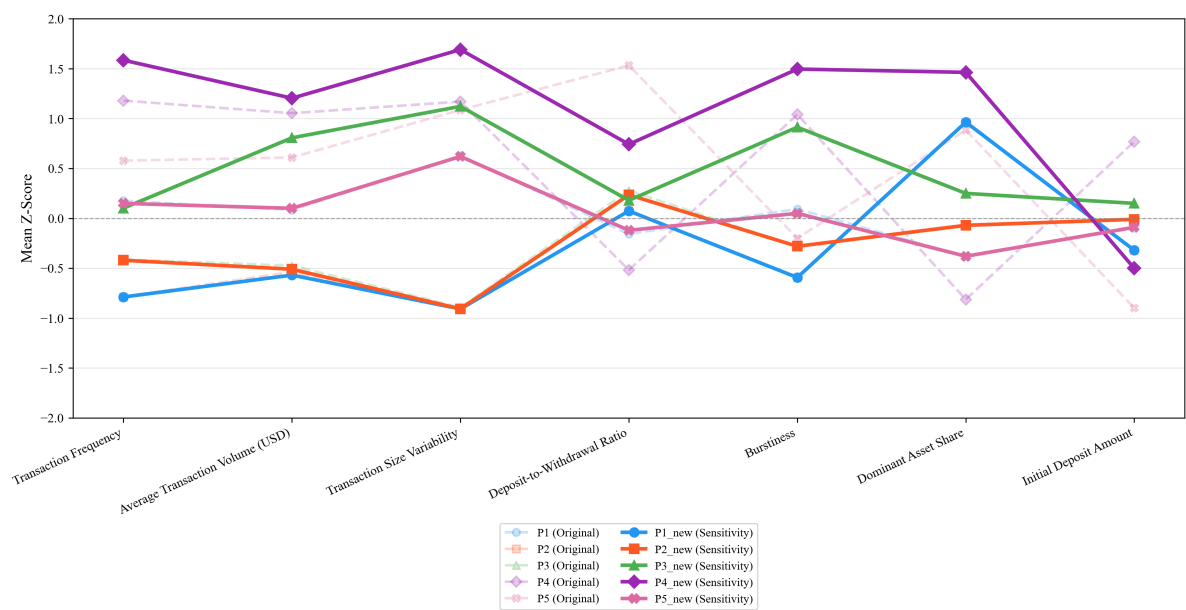


Figure B.5: Comparison of standardized indicator means between the original and internal user sensitivity analysis of the 5-profile solutions

This visual comparison of Figure B.4 indicates that the general behavioral profile structure remains relatively consistent after removing internal accounts. Deviations are mainly found for Profile 5, which shows stronger pattern deviations for burstiness, dominant asset share, and initial deposits. However, mostly line shapes of the profiles remain comparable across both model specifications. Specifically, Profile 2 and 3 remain highly comparable. However, the sensitivity analysis results in sharper profile separation and more extreme indicator values for certain profiles, suggesting that internal accounts partially make behavioral heterogeneity in the original solution appear less pronounced.

Supplements for Profile Comparison

C.1. Distributions of Indicators across Profiles

Figure C.1 provides a more detailed overview of the distribution of the indicator variables across the five latent profiles.

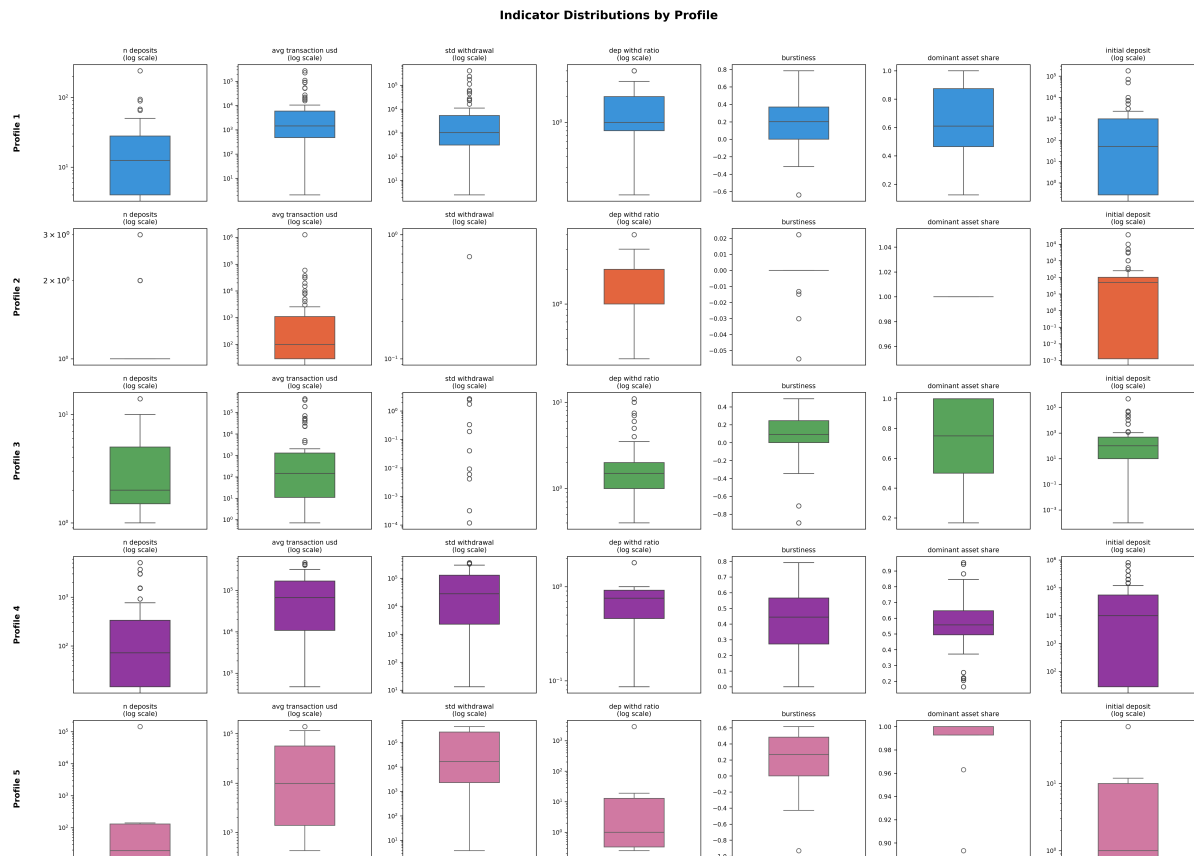


Figure C.1: Boxplots of LPA indicators across the five latent profiles
Note. Logarithmic scales were applied to the boxplots, as variables are still highly skewed.

C.2. Characteristics of Internal Users

As comparisons are often made with the presence of internal users, Table C.1 shows the spread of users identified as administrative or internal to Cryptex across the obtained behavioral profiles.

Table C.1: Distribution of selected users across latent profiles

Profile	Number of internal users	Total users	% of cluster	Observed %
1	6	76	7.9	37.5
2	2	67	3.0	12.5
3	2	59	3.4	12.5
4	5	47	10.6	31.3
5	1	9	11.1	6.3

Note. '% of cluster' indicates the proportion of users within each profile that belong to the selected subset.
'Observed %' reflects the distribution of selected users across profiles.

Post-Hoc Analysis Supplements

D.1. Address-Sharing Network Analysis

Table D.1 below provides additional descriptive statistics for the network metrics per profile, as discussed as in the main text. The table shows more detailed information on the differences in network connectivity, overlap strength, structural positioning, and component membership of users within the withdrawal-based address-sharing network. All reported metrics are based on the withdrawal network after excluding internal accounts.

Table D.1: Network centrality measures and component presence per profile

Metric	Profile	N	Mean	Median	SD	Range
Degree centrality	Profile 1	13	1.462	1.0	0.967	1–4
	Profile 4	25	3.720	3.0	3.048	1–12
Weighted degree centrality (strength)	Profile 1	13	1.923	1.0	2.216	–
	Profile 4	25	9.360	5.0	9.032	–
Weighted betweenness centrality	Profile 1	13	0.000	0.0	0.000	–
	Profile 4	25	0.062	0.0	0.113	–
Users with weighted betweenness > 0 (%)	Profile 1	13	0.0	–	–	–
	Profile 4	25	40.0	–	–	–
Component size	Profile 1	13	23.462	33.0	–	–
	Profile 4	25	30.520	33.0	–	–
Users in large component (%)	Profile 1	13	69.2	–	–	–
	Profile 4	25	92.0	–	–	–

Note. Weighted degree centrality reflects the total amount of shared withdrawal-wallet overlap across connected users. Weighted betweenness centrality was calculated using inverse edge weights as distances. Large component presence refers to the proportion of users belonging to the largest connected component in the withdrawal network.

Figure D.1 represents the graphical visualization of the withdrawal network that was used for analysis. For ethical considerations, user IDs are omitted from both figures to prevent the identification of individual users. Moreover, as explained in Chapter 8, the deposit network is considered too small for analysis and therefore neglected.

Table D.2 repeats the profile colors for correct interpretation of the visual.

Profile 1: Moderately Active Diversified Users	
Profile 2: Extreme Passive Accounts	
Profile 3: Low-Activity Users	
Profile 4: High-Volume Active Traders	
Profile 5: Systematic Accumulators	

Table D.2: Legend for profile colors used in network visualizations (corresponding with LPA profiles)

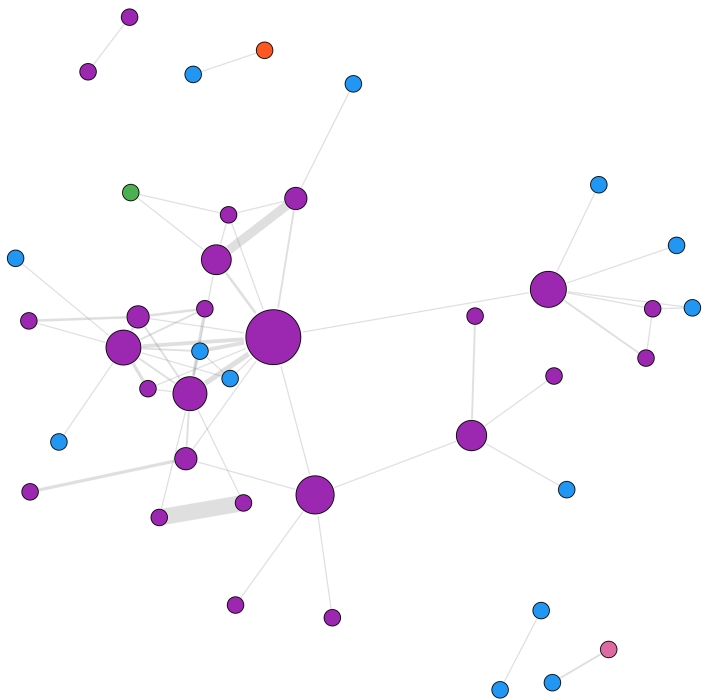


Figure D.1: Visual representation of withdrawal network based on address-sharing, without internal users

Within Figure 6.2, nodes represent users and edges represent shared withdrawal-wallet overlap between users. Here, the thickness of an edge reflects the number of unique shared withdrawal wallets, the size of the node reflects a higher weighted betweenness centrality, and node color indicates the modal class membership. For betweenness centrality calculations, edge weights were transformed into distances using the inverse of overlap strength ($distance = 1/weight$).

Based on the visual, it becomes clear that there are structural differences between the profiles regarding address-sharing patterns. Profile 4 occupies a more central position within the network and is also the only profile associated with larger node sizes. This reflects higher weighted betweenness centrality scores, which in turn suggests that Profile 4 connects different parts of the shared-wallet overlap network more often. As a result, they appear to play a more central role in linking different address-sharing structures together.

Importantly, this does not necessarily indicate direct coordination or a leadership role. Rather, it suggests that users within Profile 4 are connected to a wider range of shared withdrawal-wallet patterns compared to the other profiles. In contrast, the remaining profiles appear to be more peripheral within the network and mainly connected within small and isolated clusters.

D.2. Blockchain Coverage and Illicit Exposure

Table D.3 shows the profile-specific deviations in coverage and exposure, compared to the overall study population proportions. It provides additional insight on how the deviation plots in the main text are constructed.

Table D.3: Coverage and illicit exposure deviations across latent profiles

Profile	Profile size	Coverage (%)	Δ Coverage	Exposure (%)	Δ Exposure
1	76	36.8%	+6.6	46.4%	-3.6
2	67	9.0%	-21.3	33.3%	-16.7
3	59	22.0%	-8.2	46.2%	-3.8
4	48	61.7%	+31.5	55.2%	+5.2
5	9	22.2%	-8.0	100.0%	+50.0
Overall	30.2%	–	50.0%	–	

Note. Δ values indicate deviations in percentage points relative to the overall dataset or analyzed subset.

Within Table D.3, coverage represents the proportion of users within each profile included in the exposure analysis. The percentage represents the number of users included on the Ethereum, Blockchain or Tron blockchain, compared to the total study population. Exposure represents the proportion of users within this analyzed subset with direct illicit exposure. Of the 258 users, 78 users are present within the exposure analyses, representing the 30.2% in the Table above. Similarly, 36 users of 78 have direct illicit exposure, explaining the presented overall 50%. The exposure of Profile 5 is 100%, but should be interpreted with care, due to the size of the profile.

D.3. KYC Applications and Verifications

Below, Table D.4 provides an overview of the distribution of KYC statuses across profiles, distinguishing between the first and latest KYC-status of a user.

Table D.4: Internal compliance characteristics per latent profile

	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5
KYC verification status					
<i>First KYC status</i>					
Approved (%)	60.5	61.2	50.8	46.8	33.3
Not approved (%)	14.5	22.4	30.5	2.1	0.0
Missing (%)	25.0	16.4	18.6	51.1	66.7
<i>Latest KYC status</i>					
Approved (%)	72.4	83.6	81.4	48.9	33.3
Not approved (%)	2.6	0.0	0.0	0.0	0.0
Missing (%)	25.0	16.4	18.6	51.1	66.7

Note. Values represent percentages within each profile. “Missing” refers to users for whom no KYC status was available.

As shown in Table D.4, substantial differences can be observed between the initial and latest KYC statuses across the latent profiles. Specifically, Profiles 2 and 3 initially showed relatively higher proportions of disapproved statuses, whereas the latest recorded statuses indicate considerably higher approval rates. In contrast, Profiles 4 and 5 continue to show relatively large proportions of missing KYC information, suggesting lower overall adherence to KYC-verification procedures within these two profiles.

Table D.5 shows the input for the deviation plot in the main body of Chapter 8.

Table D.5: KYC approval and missing-status deviations across latent profiles

Profile	Profile size	Approved (%)	Δ Approved	Missing (%)	Δ Missing
1	76	72.4	+0.7	25.0	-2.5
2	67	83.6	+11.9	16.4	-11.1
3	59	81.4	+9.7	18.6	-8.9
4	47	48.9	-22.8	51.1	+23.5
5	9	33.3	-38.4	66.7	+39.1
Overall	—	71.7	—	27.5	—

Note. Δ values represent deviations in percentage points relative to the overall distribution of approved and missing KYC statuses in the dataset.

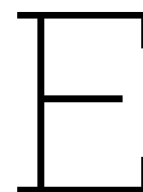
D.4. Privacy Coin Use

Table D.6 shows the input for the deviation plot presented in the main body of Chapter 8.

Table D.6: Privacy coin use deviations across latent profiles

Profile	Profile size	Privacy coin use (%)	Δ Privacy Coins
1	76	19.7	+2.7
2	67	1.5	-15.6
3	59	10.2	-6.9
4	47	44.7	+27.6
5	9	11.1	-5.9
Overall	17.0	—	—

Note. Δ values represent deviations in percentage points relative to the overall distribution of privacy coin use in the dataset.



Focus Group Protocol

E.1. Pre-Focus Group Questionnaire

This section includes the pre-focusgroup questionnaire that respondents were asked to complete in order to collect background information on their experience within law enforcement and context of the research topic. This supports the interpretation of the findings from this research part.

Participant Background

Please answer the following questions prior to the focus group session.

Role and Experience

- Please briefly describe your current role.

Experience with Cryptocurrency Investigations

- How many years of experience do you have with cryptocurrency-related investigations?

Involvement with Transaction Analysis

- How frequently do you work with transaction data?

- ☐ Daily
- ☐ Weekly
- ☐ Occasionally
- ☐ Rarely

Type of Work

- What types of tasks related to cryptocurrency investigations do you primarily perform?
(*e.g., transaction analysis, case development, intelligence, technical support*)
 - Do you specialize in a particular area within virtual financial crime? If so, please specify.
-

E.2. Focus Group Structure

This section describes the questions proposed during the focus group. Table E.1 shows which question addresses which goal of the focus group. This structure ensured that both the substantive validity of the profiles and their practical and methodological implications were systematically explored within the limited time frame of the session. The focus group was held in Dutch.

Table E.1: Alignment between research objectives and focus group questions

Research objective	Covered in
Current practices/context	Question 1
Validation of profiles	Question 2
Interpretation	Question 3
Investigative relevance	Questions 3 and 4
Limitations	Question 5
Applicability in other contexts	Question 6

E.2.1. Respondents' Introduction

Participants were provided with a hand-out prior of the session, containing information about the motivation behind the research, a description of the obtained profiles and a general explanation of the used methodology (LPA). This was done to limit time during the focus group that needed to be spent on explaining the research.

Before starting the focus group, respondents were encouraged to share any thoughts or opinions, complement or contradict one another where that was considered necessary and reminded there are no right or wrong answers. Moreover, respondents signed an Informed Consent, agreed upon and in line with guidelines established by the Human Research Ethics Committee from TU Delft. This Informed Consent document states that results will only be used for academic purposes only and includes privacy agreements.

E.2.2. Focus Group Protocol

The focus group was structured around a set of core questions, designed to elicit participants' perspectives on current investigative practices, the validity of the identified latent profiles, and their potential applicability in practice. The questions were organized into thematic blocks as shown in Table E.1, each addressing a specific aspect of the objectives of this qualitative research part, or necessary context.

1. Current Investigative Practices

- **How are datasets used in your work to identify suspicious users or transaction patterns, and what challenges or limitations do you encounter in this process?**

Optional follow-up questions:

- Do you primarily rely on transaction-level indicators, network structures, or user-level patterns?
- Where do you experience the greatest uncertainty in your analyses?

2. Validation of Latent Profiles

- **To what extent do the presented profiles reflect patterns that you recognize as realistic or representative of practice?**

Optional follow-up questions:

- Are there specific profiles that stand out as particularly plausible or implausible? Why?
- What types of users or behaviours do you associate with these profiles?

3. Interpretation in an Investigative Context

- **How would you interpret these profiles within an investigative context?**

Optional follow-up questions:

- Which profiles would you prioritize for further investigation?
- Do you perceive differences in risk or relevance across the profiles?

4. Practical Usefulness

- **In what ways, if any, could these or such profiles support your investigative work?**

Optional follow-up questions:

- At which stage(s) of an investigation would such profiles be most useful?
- How might access to these profiles influence your workflow or decision-making?

5. Critical Reflection

- **What are the main limitations or risks associated with the use of such profiles?**

Optional follow-up questions:

- In what situations could these profiles be misleading?
- Are there important aspects or data elements currently missing?

6. Broader Applicability

- **Do you see potential for applying this method, or similar approaches, in other investigative contexts or datasets?**

Optional follow-up questions:

- Could this approach be combined with other analytical methods?
- In which domains or use cases would this be most valuable?

7. Closing Question

- **Are there any additional remarks or insights you would like to share?**

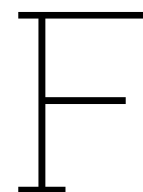
The discussion was enhanced as much as possible by the moderator by asking questions if others agreed with certain perspectives, if anyone had different experiences, or if certain statements could be further elaborated on.

E.2.3. Respondent Overview

Respondent	Position description
1	Financial detective
2	Cyber investigation specialist
3	General project leader
4	Financial detective

Table E.2: Overview of focus group respondents

All respondents were operating at the Dutch Financial Investigation and Prosecution service (FIOD). Two respondents operated within a generalized investigative team (AOT) and two respondents at a specialized team focus on financial cybercrime.



Focus Group Summary of Results

The focus group was physically conducted on Thursday April 23, 2026 at 13:45. The meeting was recorded on a secured mobile device and lasted 1:00:16. Recordings are not publicly available and were deleted after official completion of this thesis, in correspondence with the informed consent by the respondents. As respondents occasionally already contributed to subsequent questions in their answers, this led to questions being combined into a single section in this appendix chapter, which is indicated by referring to the question in the title heading.

F.1. Current Practices (Q1)

The focus group started by questioning the respondents on current investigative practices and the use of similar, large datasets. Respondents indicated that such datasets are usually available after a seizure or arrest, often when an initial case is already nearing completion. In these situations, datasets are not only used for the original investigation, but frequently for initiating new investigations and official records. Within the more specialized teams, they are sometimes revisited for subsequent analyses, particularly for the attribution of additional individuals. Additionally, these datasets may also be shared with other (foreign) investigative services.

It was mentioned by respondents from the AOT that an initial exploratory scan of the data is commonly performed. This often involves filtering on basic transactional characteristics, such as transaction volume, number of transactions, and the balance between deposits and withdrawals. These indicators are used to identify potentially relevant accounts for further investigation and to potentially request support from specialized teams. It was implicitly mentioned that a certain trade-off plays a role in this process, leading to a focus on accounts that are expected to yield the highest impact (*"we tend to focus on the big money"*), typically those that involve larger transaction volumes.

Next to this, respondents consistently referred to a set of transactional dimensions that are commonly used for initial data filtering for investigative research. Dimensions that stood out during the focus group were transaction volume, the number of transactions, the ratio and structure between deposits and withdrawals, temporal patterns, especially bursts of activity, the types of assets used (e.g., privacy coins), and whether funds are converted into fiat currency.

During the focus group, differences in roles between investigative teams frequently stood out. General investigative teams (AOT) indicated that working with highly technical datasets is not part of their primary operations. Instead, they may encounter such data more incidentally and when considered relevant, pass it on to specialized teams. Their own analytical work more often involves smaller or more practical datasets (e.g., obtained from PGP devices), which primarily aid in building a profile of a potential suspect. On the other hand, the more specialized teams engage more directly with large-scale transaction datasets and apply more advanced analytical techniques.

F.2. Interpretation and Validation of the Profiles (Q2/Q3)

The interpretation of the profiles was found to be challenging. During the discussion, no strong reactions and relatively long silences were observed, which indicated a sense of uncertainty. Respondents did not express strong agreement or disagreement with the presented profiles; rather, they indicated that the observed patterns were very generic and therefore very difficult to validate without additional context. Therefore, the profiles were generally perceived as potentially realistic representations of behavior, but not specific enough to support definitive conclusions or validation.

A recurring theme in the discussion was the importance of additional information that very much depends on the context of the investigation. Respondents emphasized that key variables used in practice, such as exposure (which can be detected through Chainalysis, also explained in Chapter 4), the conversion to fiat currency, and the type of assets used (e.g., stablecoins or privacy coins), are critical for interpretation but not fully captured in the presented profiles. In particular, the distinction between funds remaining versus moving into the fiat system was highlighted as an important indicator in investigative practice. However, during the discussion it was observed that some of these additional variables or questions raised to interpret the profiles defined in correspondence with the context in which Cryptex operated (e.g., the absence of formal KYC-procedures and similarly, the possibility to create multiple accounts).

Certain profiles triggered more discussion than others. Profile 5 was consistently identified as notable due to its extreme characteristics, but also considered unstable and potentially non-generalizable because of its small size. Profile 2 also attracted attention, specifically because of the very minimal activity (e.g., a single transaction). It was mentioned that such profiles might normally be deprioritized as investigations usually filter on volume, respondents noted that this could also indicate specific behaviors such as “throw-away” accounts that are used once and smurfing tactics, also again taking into account the context of Cryptex as a non-KYC exchange, where creating multiple accounts was relatively easy.

Profile 4 was associated with more structured or automated trading behavior by one respondent, potentially involving the use of APIs. Respondents suggested that additional data on the method of access (e.g., app versus API) could provide valuable insight into user behavior. More broadly, multiple respondents expressed a strong need for additional variables to meaningfully interpret the profiles, indicating that the current representation lacks sufficient contextual precision.

Importantly, respondents emphasized that it is too soon to assign criminal labels solely based on these profiles. While certain transactional patterns (e.g., frequent deposits and withdrawals without trading activity on a centralized exchange and without a legitimate economic rationale) could raise suspicion, meaningful interpretation requires contextual and investigative information and nuance. This was particularly stressed by one respondent regarding the use of central exchanges, where certain behaviors may resemble mixing practices but no conclusions can be drawn without further evidence.

F.3. Relevance and Applicability in Other Contexts (Q3/Q4/Q6)

During the focus group, it became clear that practical relevance of the profiles strongly depends on role and expertise. For general investigative teams (AOT), the profiles were seen as potentially useful as a filtering method or “handbook” that supports initial data exploration and understanding. In this context, profiles could help identify whether an account exhibits unusual behavior and whether it should be escalated to a specialized team. As such, the profiles may function as a form of preliminary sorting or filtering mechanism, rather than as direct evidence. However, the respondents agreed that this should be first empirically generalized across exchanges to create such handbooks and such grouping of profiles would not be used in official reports.

For more specialized investigators, the direct operational value was considered more limited. One respondent explicitly noted that the profiles would not be used as steering information, but rather as a secondary check: given an account of interest, the profile could be used to assess whether the observed behavior aligns with expectations or deviates from typical patterns.

Additionally, the profiles were considered potentially useful for testing specific hypothesis on criminal behavioral patterns and a following comparative analysis. For example, once a specific modus operandi has been identified, profiles could help identify other accounts that show similar behavior. This suggests a possible application in pattern recognition and clustering of related accounts within large datasets.

The discussion also indicated that the method may contribute to a broader analytical perspective. Respondents noted that investigative practices tend to focus on high-volume accounts, which may lead to the underexploration of alternative behavioral patterns (e.g., low-volume or smurfing strategies). In this sense, the profiles may support awareness of less obvious patterns, or other *modus operandi* that become visible within the data, which is in the context of this research interpreted as "detection bias".

Finally, as mentioned before, respondents expressed interest in the potential generalizability of the approach. While some expected similar patterns to emerge across different exchanges, others noted that platform-specific characteristics (such as the absence of KYC procedures) may significantly influence user behavior.

This highlights the importance of contextualizing the profiles within the specific environment from which the data is derived and raises the question how easily such handbooks are created if context is so important.

Overall, the discussion highlighted that the interpretation and usefulness of the profiles are highly dependent on the intended purpose. While the profiles may offer value for exploratory analysis and segmentation, the application in attribution of individuals remains very limited without integrating context specific, additional data points and in addition human-intel, which is considered to bring a certain nuance and specific steering that cannot be obtained through data alone.

F.4. Limitations and Missing Elements (Q6)

A key limitation that was mentioned by respondents is the generality of the profiles. It is considered that this results in a loss of nuance and detail that is highly essential for investigative interpretation. When asked, one respondent with a specialized background indicated that additional variables could potentially improve interpretability, but also acknowledged the inherent trade-off between model complexity and statistical robustness, particularly in smaller datasets.

Several specific missing elements were identified. Most notably, exposure to other illicit funds on the blockchain (i.e., analysis through Chainalysis) was repeatedly mentioned as a crucial indicator in practice. Similarly, which already became apparent in previous questions, information on fiat conversion, asset types, and transaction context were considered highly relevant but not sufficiently represented in the current analysis.

Respondents also raised the importance of detecting specific criminal behaviors, such as smurfing, and questioned whether such patterns can be captured within the current framework. Additionally, the ability to link multiple accounts to a single actor was identified as a potentially valuable extension, particularly in the context of exchanges where account creation is unrestricted.