

FernUniversität in Hagen
Department of Mathematics and Computer Science

Spectral Analysis of Financial Crime Patterns in Transaction Networks

An Application of Persistent Laplacian Theory to Synthetic AML Data

Master's Thesis
in the Master's Degree Programme M.Sc. Mathematics

submitted by
Katharina Blaschka
Matriculation Number: 7903979

Supervisor: Prof. Dr. Delio Mugnolo
Co-Supervisor: PD Dr. Joachim Kerner

Date of Submission: August 11, 2026

Contents

1	Introduction	1
1.1	Problem Statement and Motivation	1
1.2	Mathematical Approach and Research Questions	2
1.3	Relevance	3
1.4	Scope and Limitations	4
1.5	Structure of the Thesis	5
2	Financial Crime and Structural Characteristics of Transaction Networks	7
2.1	Terminology and Regulatory Framework	7
2.1.1	Financial Crime and AML Functions	7
2.1.2	Core AML Processes: KYC, TMS, SARs	8
2.1.3	Regulatory Requirements and Supervisory Expectations	9
2.2	Money Laundering in Practice: Typologies and Patterns	10
2.2.1	The Money Laundering Cycle: Placement, Layering, Integration	10
2.2.2	Common Typologies	11
2.2.3	Graph-Based Views on Transaction Networks	12
2.3	Graph-Theoretic Patterns in Transaction Networks	12
2.3.1	Cycles, Rings, and Circular Trade	13
2.3.2	Fan-In and Fan-Out Structures (Smurfing)	14
2.3.3	Layering Cascades	14
2.3.4	Strongly Connected Components and Circulation	15
2.3.5	Temporal Bursts and Dynamic Graphs	16
2.3.6	Bridging the Gap: From Directed Flows to Symmetrised Topological Shapes	17
2.4	Existing Approaches and the Case for Spectral Methods	18
2.4.1	Rule-Based Approaches	18
2.4.2	Machine Learning and Graph-based Methods and Remaining Gaps	18
2.4.3	Why Spectral and Topological Methods?	19

3	Graph, Spectral, and Topological Foundations	21
3.1	Graphs for Transaction Networks	22
3.1.1	Directed graphs for transactions	22
3.1.2	Multi-edge and weighted models	23
3.1.3	Local structure: subgraphs and strongly connected regions	26
3.2	Classical Graph Laplacians and Spectra	26
3.2.1	Adjacency and Degree matrices	28
3.2.2	Graph Laplacians	30
3.2.3	Basic spectral properties	33
3.3	Paths, Connectivity, and Cycles	36
3.3.1	Walks, Paths, and Cycles	37
3.3.2	Connectivity and Components	38
3.3.3	Spectral Indicators of Connectivity	39
3.4	Simplicial Complexes and Boundary Operators	40
3.4.1	Simplices and Simplicial Complexes	41
3.4.2	Chain Groups	42
3.4.3	Boundary Operators	43
3.4.4	Homology and Betti Numbers	44
3.5	The Hodge Laplacian	45
3.5.1	Chain and cochain spaces, inner products, and adjoints	46
3.5.2	Definition of the Hodge k -Laplacian	48
3.5.3	Hodge k -Laplacians on cochains	48
3.5.4	Harmonic Cochains and Hodge Decomposition	49
3.5.5	Betti Numbers and Spectra	50
3.6	Persistent Laplacians	51
3.6.1	Motivation: Limitations of Persistent Homology	52
3.6.2	Filtrations in Practice	53
3.6.3	Persistent Homology Groups and Betti Numbers	53
3.6.4	Definition of Persistent Laplacians	55
3.6.5	Matrix Representation and Spectral Information	55
3.6.6	Example: Transaction Ring in a Filtration	56
4	Persistent Laplacians for Transaction Network Analysis	57
4.1	From Persistent Homology to the Persistent Laplacian	57
4.1.1	Three Operators, Three Limitations	57
4.1.2	Operator Structure and Key Properties	60
4.2	Definition and Fundamental Properties	62
4.2.1	Definition via the Restricted Boundary Operator	62
4.2.2	Self-Adjointness and Positive Semidefiniteness	64
4.2.3	The Betti–nullity correspondence	65

4.3	Schur-Complement Representation and Algorithms	69
4.3.1	Block Structure of the Boundary Matrix	69
4.3.2	The Schur-Complement Formula	70
4.3.3	Algorithm 1: Direct Construction	72
4.3.4	Algorithm 2: Via the Schur complement	73
4.4	Spectral Properties and Stability	74
4.4.1	Eigenvalue Monotonicity	74
4.4.2	Algebraic Stability	77
4.4.3	Persistent Cheeger Inequality	79
4.5	Persistent Laplacians on Non-Branching Complexes	81
4.5.1	Motivation: Transaction Chains	81
4.5.2	Spectrum of the Path Gram Matrix	81
4.5.3	Explicit Spectra for Path Inclusions	82
4.6	Spectral Barcodes and Information-Theoretic Quantities	86
4.6.1	Classical Barcodes and Their Spectral Extension	86
4.6.2	Spectral Entropy	86
4.7	Filtration Construction from Transaction Networks	89
4.7.1	Data Basis: The AMLworld Transaction Graph	89
4.7.2	Temporal Filtration via Sliding Windows	90
4.7.3	Volume-Based Filtration	91
4.8	Spectral Features for AML Detection	92
4.8.1	Window-Level Feature Definitions	92
4.8.2	Mapping to AML Patterns	93
4.8.3	Account-Level Aggregation	93
4.8.4	Algorithmic Pipeline for Chapter 5	93
5	Results and Discussion	97
5.1	Structure of this chapter	97
5.2	Data preparation and graph construction	98
5.2.1	Identification of the illicit vertex set V_{ill}	98
5.2.2	Fraud-role annotation	99
5.2.3	Loading transactions and removal of self-loops	99
5.2.4	One-hop filter and induced subgraph	99
5.2.5	Two-hop extension to target budget	99
5.2.6	Final induced subgraph topology	100
5.2.7	Feature engineering and currency conversion	100
5.2.8	Fraud labels on nodes versus edges	101
5.3	Construction of the transaction graph	102
5.3.1	Directed weighted multigraph D	102
5.3.2	Descriptive graph statistics	102

5.3.3	Strongly connected components	104
5.3.4	Exploratory label analysis	104
5.3.5	Persisted output files	107
5.4	Simplicial complex, filtrations, and homology	107
5.4.1	Graph preparation and symmetrisation	107
5.4.2	Clique complex $K(G_{\text{work}})$	109
5.4.3	Volume-based filtration	110
5.4.4	Persistent homology via GUDHI	110
5.4.5	Temporal filtration and burst indicator	111
5.4.6	Persisted output files	112
5.5	Persistent Laplacians and spectral features	112
5.5.1	Filtration pairs and Schur-complement algorithm	112
5.5.2	Betti-nullity correspondence: empirical validation	114
5.5.3	AML feature vector $f^{(s)}$	115
5.5.4	Empirical results: volume-based filtration	116
5.5.5	Temporal filtration and burst detection	118
5.5.6	Spectral gaps and path calibration	119
5.5.7	Monotonicity and stability verification	122
5.6	Discussion of Empirical and Account-Level Results	122
5.6.1	Birth-step mapping: correction and calibration	122
5.6.2	Account-level aggregation	123
5.6.3	Feature Effects and Fraud Typologies	126
5.6.4	Scope and Structural Limitations	127
5.7	Summary of Research Questions	128

Chapter 1

Introduction

1.1 Problem Statement and Motivation

Money laundering is estimated to account for 2 to 5 percent of global gross domestic product, corresponding to approximately 800 billion to 2 trillion US dollars annually [Bakhshinejad et al., 2024, Section 1]. Beyond its economic scale, it constitutes an enabling infrastructure for organised crime: 76 percent of the most dangerous criminal networks identified in the 2024 Europol report are active across two to seven countries [FATF et al., 2025, p. 19], confirming that the problem is inherently transnational and structurally complex.

Financial institutions are required by law to detect and report suspicious transaction activity. The primary instrument for this is the Transaction Monitoring System (TMS), which evaluates incoming transactions against predefined rule sets and generates alerts for human review. Despite decades of refinement, industry-standard TMS produce a false alert rate (FAR) of 95 %–98 % [Bakhshinejad et al., 2024, § 1]: nearly every alert generated is a false positive. Columbia School of International and Public Affairs (SIPA) [2018] confirm that 90 %–95 % of alerts in typical deployments require no further action, producing enormous investigative overhead while diverting attention from genuinely suspicious activity. The ratio of suspicious to non-suspicious transactions in Anti Money Laundering (AML) datasets ranges from 1:100 to 1:1,000 [Bakhshinejad et al., 2024, § 1.2.2], making the detection problem one of extreme class imbalance.

A central limitation of rule-based approaches is that they primarily evaluate scalar properties of individual transactions. Many money-laundering typologies, however, manifest themselves through relational and temporal patterns involving several accounts and transactions. Such patterns may include cyclic flows, fan-in/fan-out structures, layering chains, or abrupt changes in local network topology. Detecting them therefore requires methods that explicitly represent the network structure of transaction data.

This does not imply that every individual transaction carries a separable structural signal. Rather, the

relevant information may emerge only after aggregation across several transactions, filtration levels, or accounts. The present thesis investigates to what extent spectral and topological quantities can provide a mathematically interpretable description of such multi-scale structures.

1.2 Mathematical Approach and Research Questions

This thesis investigates money laundering from a mathematical perspective. Transaction data are modelled as finite directed weighted multigraphs $D = (V, A, w)$, in which vertices represent accounts, arcs represent transactions, and weights encode transaction amounts. The central question is:

How can spectral and topological operators, in particular persistent Laplacians, be used to characterise money-laundering patterns in transaction networks, and at which level of aggregation, individual transactions, filtration windows, or accounts, do these operators provide usable structural information?

This primary question decomposes into three secondary questions:

1. How can standard AML typologies, such as cycles, fan-in/fan-out structures, layering cascades, strongly connected components, and temporal bursts, be formalised as graph-theoretic and topological objects, and how can classical, Hodge, and persistent Laplacians be constructed on transaction graphs to encode these structures spectrally?
2. Which theoretical properties of persistent Laplacians, including the kernel–homology correspondence, spectral formulae on elementary substructures such as paths, and the monotonicity of the up-Laplacian spectrum under filtration extension, can be assessed numerically on a concrete filtered transaction graph? In particular, under which conditions does the coarse discretisation required for tractable computation on realistic transaction volumes limit the numerical observability of these theoretical guarantees?
3. How do the resulting spectral features behave on synthetic AML data at different levels of aggregation, individual transaction types, filtration windows, and accounts? Do fraud-related structures become distinguishable from benign background activity at any of these levels, and what role do threshold calibration and the choice of aggregation scheme play in this outcome?

Central working hypothesis. *We hypothesise that the kernel of the persistent Laplacian $\Delta_q^{K,L}$ recovers persistent Betti numbers exactly on a filtered transaction graph, and that its non-zero spectrum is consistent with the spectral behaviour derived for elementary path-like substructures. We further hypothesise that individual transactions, viewed in isolation, carry too little structural information to yield a stable descriptive separation under extreme class imbalance. By contrast, aggregating spectral contributions over the transactions incident to an account may produce a descriptive shift towards higher feature values*

for fraud-labelled accounts than for background accounts.

This latter pattern is not interpreted as evidence of a validated classification method. Its practical relevance depends on the consistency between the filtration thresholds, the graph on which they are calibrated, and the volume and transaction-frequency distribution of the population to which the resulting features are transferred.

The approach proceeds in three steps. First, transaction networks are modelled as directed weighted multigraphs and analysed using directed, combinatorial, and symmetrised Laplacians. Second, for the topological and persistent-spectral analysis, the directed graph is mapped to a symmetrised weighted graph and lifted to a simplicial complex via the clique construction. This deliberately abstracts from the orientation of individual transactions; directed flow information is therefore not recovered by the topological construction, but is retained separately for the combinatorial analysis of directed structures such as strongly connected components and in-/out-degree asymmetries.

The clique construction enables the definition of Hodge Laplacians, whose kernels encode independent one-dimensional homological features through the first Betti number β_1 . Third, a volumetric filtration of simplicial complexes, indexed by decreasing transaction-amount thresholds, is constructed on a pattern-centred working subgraph. Persistent Laplacians $\Delta_q^{K,L}$ are then used to extract spectral features along this filtration. These quantities are mapped descriptively to transaction types and aggregated at the account level, where they are compared with the fraud labels of the synthetic AMLWORLD dataset of [Altman et al. \[2024\]](#).

1.3 Relevance

Mathematical relevance. Transaction networks provide a structured setting for the application and further development of modern spectral-topological methods. The persistent Laplacian, introduced and studied by [Wei and Wei \[2024\]](#) and [Mémoli et al. \[2022\]](#), refines persistent homology by recovering persistent Betti numbers from the kernel of a self-adjoint operator while encoding additional geometric information in its non-zero spectrum. Applied to simplicial filtrations derived from transaction data, this framework connects spectral graph theory, Hodge theory on simplicial complexes, and persistent homology in a unified operator-theoretic setting. The directed nature of transaction graphs further motivates the Laplacian theory for digraphs developed by [Li and Zhang \[2012\]](#). In this thesis, directed operators are used to retain and analyse flow-related information, whereas the persistent-spectral and topological constructions are defined on a symmetrised graph. This separation makes explicit that the latter constructions describe undirected structural properties of the transaction network rather than recovering the original orientation of individual transactions; the corresponding framework is developed in Chapters 3 and 4. The exact kernel identities and the spectral formulae derived for elementary substructures are mathematically well behaved at the analytical level. Their numerical evaluation on large, coarse, and discretised transaction filtrations may nevertheless be sensitive to the selected threshold sequence, matrix conditioning, and floating-point implementation. This

distinction between analytical validity and numerical observability motivates further work on stable persistent Laplacian computations for large transaction networks.

Applied and regulatory relevance. Under Article 8(2) of Directive (EU) 2015/849 [European Parliament and Council \[2015\]](#), obliged entities must document and justify their risk assessment methodology. This creates a practical need for methods whose quantities, modelling choices, and limitations can be communicated clearly within a transaction-monitoring process (see Subsection [2.1.2](#)).

Spectral invariants such as the Fiedler value $\lambda_1(L)$, the first Betti number β_1 , and the persistent spectral gap $\lambda_{\text{gap}}^{(1)}$ have explicit mathematical definitions and can be related to properties such as connectivity, cyclic structure, and bottlenecks. They may therefore complement less-transparent representations in exploratory AML analyses. However, interpretability alone does not establish regulatory suitability, operational effectiveness, or a reduction in false alerts. Such claims would require validation on representative data, an explicitly defined decision rule, and a comparison with relevant operational baselines.

1.4 Scope and Limitations

This work has a mathematical focus: The emphasis is on the derivation, proof, and spectral-theoretical interpretation of the relevant operators, not on the development of a fully functional AML system. Accordingly, the following limitations apply:

- The empirical analysis uses the synthetic AMLWORLD dataset [Altman et al. \[2024\]](#). No real production transaction data are used.
- No large-scale performance benchmarking against industrial TMS is performed, and no classification metric beyond descriptive group comparisons is introduced, in order to restrict all empirical claims to quantities directly derived from the theory of Chapters [3](#) and [4](#).
- The simplicial complex construction is restricted to clique complexes of dimension at most two (vertices, edges, triangles), which suffices for the 0- and 1-Hodge and persistent Laplacians of primary interest.
- For computational tractability, the persistent Laplacian is evaluated on a pattern-centred working subgraph of moderate size (Definition [5.4.1](#)) rather than on the full transaction network. The resulting filtration thresholds are calibrated on this subgraph, and their transfer to the full account population is examined critically in Chapter [5](#) rather than assumed to be unproblematic.
- Scalability considerations are addressed at the algorithmic level (Chapter [4](#)) but not empirically validated at production scale.

- The transfer of filtration thresholds from the working graph to the full account population is not a recalibration on the full edge distribution. Consequently, account-level comparisons are interpreted as descriptive results under a specific thresholding and aggregation scheme, rather than as estimates of out-of-sample detection performance.

1.5 Structure of the Thesis

The thesis is organised as follows.

Chapter 2 provides the criminological and regulatory background. It introduces the legal definition of money laundering, the three-phase placement–layering–integration cycle, the main typologies (smurfing, TBML, shell companies, hawala), and the existing detection landscape. Each typology is described in terms of its structural signature in the transaction graph, preparing the mathematical formalisation of Chapters 3 and 4.

Chapter 3 develops the mathematical foundations. Transaction networks are introduced as directed weighted multigraphs. Classical and symmetrised Laplacians are defined and their spectral properties established. The chapter then develops the topological framework: simplicial complexes and clique complexes, boundary operators, Hodge Laplacians, and the Hodge decomposition of edge flows. It closes with an introduction to persistent Laplacians and filtrations, which are studied in full in Chapter 4.

Chapter 4 develops the theory and algorithmics of persistent Laplacians in full. The Betti–nullity correspondence, the Schur-complement representation, two explicit algorithms, spectral stability results, the persistent Cheeger inequality, and closed-form eigenvalue formulae for path inclusions are proved.

Chapter 5 presents the empirical case study based on the synthetic AMLWORLD dataset of Altman et al. [2024]. It proceeds in five stages:

- **Data basis.** Description and preprocessing of the AMLWORLD data, including the labelling structure and basic volume statistics.
- **Graph construction.** Construction of the directed weighted transaction graph and of a pattern-centred working graph used for the filtration.
- **Filtrations.** Definition of a volumetric filtration on the working graph.
- **Window-level analysis.** Evaluation of persistent Laplacians on successive filtration pairs. This analysis numerically validates the Betti–nullity correspondence and compares observed spectra with the path-spectrum formula on elementary substructures. It also examines the numerical

observability of up-Laplacian monotonicity under filtration extension, including deviations that arise under the coarse discretisation used in the empirical setting.

- **Account-level aggregation.** Mapping of spectral quantities back to individual transactions and aggregation at the account level, followed by a descriptive comparison between fraud-labelled and background accounts and a critical discussion of threshold calibration and transfer to the full account population.

Chapter 2

Financial Crime and Structural Characteristics of Transaction Networks

2.1 Terminology and Regulatory Framework

Money laundering creates a threat to financial systems' safety across the globe. According to [Bakhshinejad et al., 2024, p. 496], the total size of the money laundering activity is believed to range from 2 to 5 percent of the world's gross domestic product, which translates into 800 billion to 2 trillion US dollars. This figure makes money laundering the third-largest segment of organized crime [Deprez et al., 2025b, p. 1].

2.1.1 Financial Crime and AML Functions

Within the legal framework of the European Union, the money laundering act is defined in Article 1(3) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Article 1, (3)] as the conversion, transfer, concealment, acquisition, possession or use of property obtained through criminal activities along with the participation in, or assistance to, these activities. Criminal law minimum harmonisation in the Union is guaranteed by Directive (EU) 2018/1673 [European Parliament and Council, 2018b, (11)], which in addition to the aforementioned also prohibits self-laundering and sets a maximum period of imprisonment of four years. In the context of German law, Section 1(1) GwG [Federal Republic of Germany, 2017, § 1(1)] links money laundering directly to the criminal offence defined in Section 261 StGB [Federal Ministry of Justice, 2024, Section 261].

Money laundering has become an international crime phenomenon as a result of the processes of digitalisation and globalisation [FATF et al., 2025, p. 6]. According to the Europol study *Decoding the EU's most threatening criminal networks* (2024), 76 percent of the most dangerous criminal net-

works are active in two to seven countries [FATF et al., 2025, p. 19]. It proves that purely national rules-based frameworks are not adequate. To combat money laundering, a framework of laws and procedures (anti-money laundering (AML)) has been established, comprising customer due diligence, transaction monitoring, and suspicious activity reporting [Bakhshinejad et al., 2024, p. 496].

The concept of a *transaction* as a central operation is of crucial importance for the graph-theoretic model developed in Chapter 3. According to the German Anti-Money Laundering Act (GwG), a transaction is defined as a single act or, where a connection between them appears to exist, several acts that are intended to effect or do effect a movement of funds or another transfer of assets [Federal Republic of Germany, 2017, § 1(5)]. This definition explicitly encompasses both individual and related sequences of acts, reflecting the possibility that economically connected activities may constitute a single transaction. A graph-theoretic representation of this concept is provided in Definition 3.1.1.

2.1.2 Core AML Processes: KYC, TMS, SARs

The three core processes KYC (Know Your Customer), TMS (Transaction Monitoring Systems), and SARs (Suspicious Activity Reports) are the institutional environment in which the network detection methods introduced in Chapters 4 and 5 have to be implemented.

Know Your Customer (KYC). Under Article 13(1) of the Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 13(1)], the obliged entity is required to perform customer identification, verification of the identity of the customer, determination of the beneficial owner, evaluation of the purpose of the business relationship, and continuous monitoring of transactions. Under the GwG in Germany, there is an individual obligation to carry out continuous transaction monitoring on the basis of the risk profile of every single customer [Federal Republic of Germany, 2017, § 10(1) No. 5].

Transaction Monitoring Systems (TMS). TMS form the basis of alert generation systems for AML compliance programs [Columbia School of International and Public Affairs (SIPA), 2018]. They analyze transaction flows according to pre-defined scenarios and generate automatic alerts on possible money laundering activities. Although TMS play a critical role in the process, industry standard solutions of TMS are marked with a false alert rate (FAR) of 95 %–98 % [Bakhshinejad et al., 2024, p. 497], Columbia School of International and Public Affairs (SIPA) [2018] report an even higher percentage of 90 %–95 %. The ratio of suspicious to non-suspicious samples is one to 100 to one to 1,000 [Bakhshinejad et al., 2024, p. 497]. The imbalance calls for algorithms that focus more on the structure of anomalies than the total amount of transactions. The independence of the spectral features discussed in Chapters 3 and 4 from the transaction volume is investigated in Chapter 5.

The alert process, which includes triaging alerts, reviewing cases, and reporting to regulators [Bakhshinejad et al., 2024, p. 496], ends with the money laundering reporting officer (MLRO) who, under Section 7

GwG [Federal Republic of Germany, 2017, § 7], is directly accountable to management. The concrete analytical methodology within the TMS remains largely at the discretion of the obliged entity, creating scope for graph-based and spectral approaches.

Suspicious Activity Reports (SARs). The obligation to file SARs is anchored in Article 33 of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 33] and implemented in Germany via Section 43(1) GwG [Federal Republic of Germany, 2017, § 43(1)].

2.1.3 Regulatory Requirements and Supervisory Expectations

The main factors that govern the European Union’s AML structure are the Fourth Money-Laundering Directive [European Parliament and Council, 2015] and the Fifth Anti-Money-Laundering Directive, which covers virtual assets [European Parliament and Council, 2018a]. The main principle that guides this structure is the risk-based approach: according to Article 8(1) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 8(1)], the risk-based approach entails identifying and assessing risks related to money laundering for customers, products, regions, and transactions. In addition, Article 8(2) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 8(2)] states that entities must document their risk assessments and revise them on a regular basis. At the national level, BaFin exercises supervisory responsibility over banks under Sections 50 and 51 GwG [Federal Republic of Germany, 2017, §§ 50, 51].

Spectral and graph-based methods used inside a TMS must then comply with three regulatory dimensions based on a risk-based approach:

1. **Risk-sensitivity:** The detection rule must adapt to structural anomalies in the transaction network independent of the actual amounts transacted in compliance with Article 8(1) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 8(1)].
2. **Effectiveness:** The technique must demonstrably decrease the number of false alerts in comparison to a rule-based TMS in compliance with the supervisory expectations of BaFin under Sections 50–51 GwG [Federal Republic of Germany, 2017, §§ 50, 51].
3. **Interpretability:** Detection rules must be mathematically defined and explainable for an obliged entity to explain its risk assessment approach in Article 8(2) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 8(2)].

Whether the spectral operators introduced in Chapters 3 and 4 satisfy these three dimensions is assessed in Chapter 5.

2.2 Money Laundering in Practice: Typologies and Patterns

Before the mathematical tools developed in Chapters 3 and 4 can be applied to transaction networks, we must first understand which structural patterns typically arise in such data. This section therefore reviews the criminological and regulatory literature on the most common forms of money laundering and uses these empirical findings to motivate the graph-theoretic and spectral formalisations introduced in the subsequent chapters.

2.2.1 The Money Laundering Cycle: Placement, Layering, Integration

This subsection is based on the framework proposed in [Association of Certified Fraud Examiners \(ACFE\) \[2025a\]](#).

Money laundering is not a single event, but rather a structured, multi-stage process aimed at concealing the illegal origin of funds and reintroducing them into the legal economic cycle. The process is typically divided into three successive phases: placement, layering, and integration. Each of these phases leaves characteristic structural traces in transaction networks, which are formalised in Chapter 4 as spectral and topological invariants.

Placement. The first stage, *placement*, concerns the physical introduction of illegally obtained cash into the financial system, typically through deposits, currency exchanges, or cross-border transfers. From a regulatory perspective, this phase is both the most critical and the most traceable, as the link between criminal proceeds and financial transactions is most direct at this point [[Association of Certified Fraud Examiners \(ACFE\), 2025a](#), pp. 2–3]. In graph-theoretic terms, the concentration of cash inflows across multiple accounts gives rise to nodes with unusually high in-degree, which can serve as indicator of anomalous activity. The formal definition of in-degree is provided in Definition 3.1.6.

Layering. If the placement stage remains undetected, the process enters its second phase, *layering*. In this phase, the aim is to disrupt the audit trail by routing funds through a chain of increasingly complex transactions spanning multiple accounts, currencies, and jurisdictions [[Association of Certified Fraud Examiners \(ACFE\), 2025a](#), p. 3]. Such repeated movements of funds between accounts naturally give rise to cyclic substructures in the transaction graph. These cycles admit a spectral characterisation via the nullity of the 1-Hodge Laplacian Δ_1 , whose kernel dimension coincides with the first Betti number β_1 . Both quantities are introduced formally in Sections 3.4.4 and 3.5.5, respectively.

Integration. The third stage, *integration*, closes the cycle by channeling the layered funds back into seemingly legitimate economic activity. At this point, detection becomes considerably more challenging, as the illicit funds now exhibit the outward characteristics of lawful income [Association of Certified Fraud Examiners (ACFE), 2025a, pp. 3–4]. From a graph-theoretic viewpoint, integration manifests itself as an expansion of the largest weakly connected component of the transaction graph, reflecting the merger of illicit flows with pre-existing legitimate structures. The resulting spectral dynamics are exploited in Chapter 4 for the retrospective identification of subgraphs associated with money-laundering activity.

2.2.2 Common Typologies

The typologies introduced below can be viewed as refinements of the three-phase cycle outlined above. Each of them gives rise to a characteristic structural signature in the transaction graph, and these correspondences are summarised in Table 2.1.

Smurfing and Structuring. One of the most widely used placement techniques is to break up large amounts of illicit cash into many smaller transactions, each carefully calibrated to remain below the thresholds that trigger reporting obligations [Association of Certified Fraud Examiners (ACFE), 2025a, p. 3]. This practice gives rise to a characteristic star-shaped subgraph: numerous low-weight source nodes funnel funds into one or a small number of target accounts. Such structures are marked by high fan-in degree and pronounced temporal bursts, both of which are formalised as detection features in Sections 2.3.2 and 4.8.

Shell Companies. Corporate vehicles that lack substantive economic activity play a central role in the layering phase [Association of Certified Fraud Examiners (ACFE), 2025b, p. 8]. Chains of such shell companies give rise to directed paths of considerable depth, often accompanied by feedback loops. These cyclic structures increase the first Betti number β_1 of the associated simplicial complex, so that elevated values of β_1 within a subgraph can serve as a topological indicator of layering activity, as formalised in Section 3.4.4.

Alternative Remittance Systems. Informal value-transfer systems such as hawala operate entirely outside the formal banking infrastructure [Association of Certified Fraud Examiners (ACFE), 2025b, p. 24]. Their bilateral settlement mechanisms leave no direct trace in conventional financial records, effectively bypassing standard transaction monitoring. From a graph-theoretic viewpoint, this gives rise to connectivity gaps that lower the algebraic connectivity λ_1 of the reconstructed transaction

graph, the corresponding spectral interpretation is developed in Section 3.3.

Bulk Cash Smuggling. Despite the increasing sophistication of the typologies discussed above, bulk cash smuggling — the physical transportation of large amounts of cash across borders — remains widespread [Association of Certified Fraud Examiners (ACFE), 2025b, p. 14]. The subsequent placement of smuggled cash leads to abrupt activations of previously dormant nodes, producing temporal bursts that can be detected via the persistent Laplacian feature $F_\beta(\tau)$ introduced in Definition 4.8.1.

2.2.3 Graph-Based Views on Transaction Networks

The global financial system is, in essence, a network: institutions, accounts, and transactions are linked through relationships that differ in volume, currency, and instrument type. Building on this perspective, Granados and Vargas [2022] show empirically that analysing the structural organisation of a financial network allows one to identify suspicious groups of agents and to characterise the ways in which their members are separated from one another [Granados and Vargas, 2022, p. 2].

This viewpoint directly motivates the spectral indicators introduced in Chapter 3. The algebraic connectivity λ_1 (see Definition 3.3.10) measures how tightly a subgraph is integrated into the surrounding network [von Luxburg, 2007, pp. 4–6][Spielman, 2025, p. 30], while strongly connected components isolate groups of accounts within which funds can circulate along closed loops [Granados and Vargas, 2022, pp. 8–9].

Weber et al. [2018] establish graph analytics as an increasingly important instrument in AML practice, noting that a natural representation of transactional data is a bipartite directed graph with account identifiers on one side and transactions on the other [Weber et al., 2018, p. 3]. Real-world AML transaction graphs, however, contain billions of edges between millions of nodes, which imposes stringent scalability requirements on any spectral method. These constraints guide the algorithmic design in Chapter 5.

These observations motivate the graph model introduced in Chapter 3, which captures directionality, edge weights, and multi-edge structures in transaction networks.

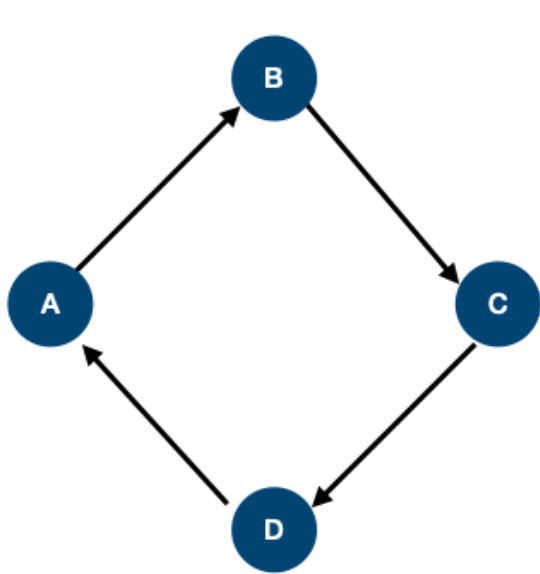
2.3 Graph-Theoretic Patterns in Transaction Networks

Transaction networks form the empirical starting point for the spectral analysis of money-laundering structures. This section describes the patterns observed in criminological practice in graph-theoretic

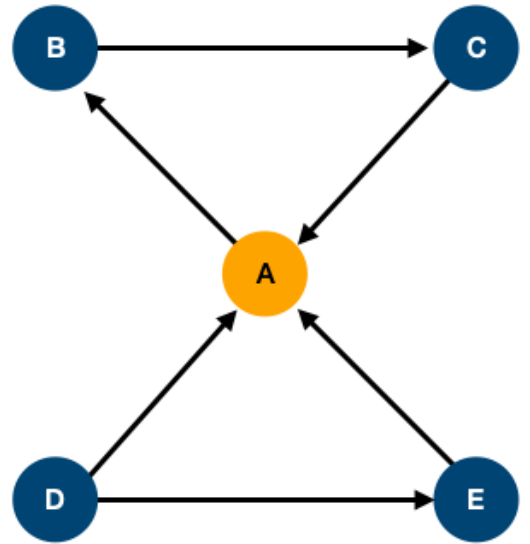
language. Each pattern corresponds to a spectral or topological signal that is formalised and quantified in Chapters 3 and 4. Table 2.1 summarises these correspondences.

2.3.1 Cycles, Rings, and Circular Trade

The most elementary pattern in criminal transaction networks is the *circulation structure*: funds travel in a closed loop that eventually returns to its starting point. Granados and Vargas [2022] formalise this as an explicit graph rule: launderers resort to deposit triangulation between the same agents in order to complete a cycle [Granados and Vargas, 2022, p. 8], and at least part of the money involved is expected to return to the originating agent [Granados and Vargas, 2022, p. 9]. Cycles of length $\ell \geq 3$ through a common node are regarded as suspicious. The detection task then reduces to finding maximal cliques in the reduced network [Granados and Vargas, 2022, p. 9].



(a) A directed cycle of length $\ell = 4$: funds originating at node *A* circulate through intermediary nodes *B*, *C*, *D* before returning to *A*.



(b) Two crossing cycles sharing a common bifurcation node *A* (highlighted): the co-occurrence of multiple cycles through the same node is the primary suspiciousness indicator.

Figure 2.1: Directed cycle structures as layering patterns. Based on [Granados and Vargas, 2022, p. 9].

As illustrated in Figure 2.1, panel (a) shows the minimal circulation structure satisfying Rule 5 of Granados and Vargas [2022]: among all possible paths in the network, closed paths are more relevant than simple paths [Granados and Vargas, 2022, p. 9]. Panel (b) extends this to the strategically more suspicious case of crossing cycles sharing a bifurcation node, which forms the basis of the clique-

reduction strategy developed in Granados and Vargas [2022].

Bonato et al. [2024] corroborate this empirically: applying a cycle-detection algorithm to unreported transactions, they identify 183 directed cycles of unreported transfers among 646 accounts [Bonato et al., 2024, p. 8]. The spectral and topological formalisation of cycles, via the first Betti number β_1 (Section 3.4.4) and the Fiedler value $\lambda_1(L^{\text{sym}})$ of the symmetrised Laplacian (Section 3.3), is given in those sections respectively.

2.3.2 Fan-In and Fan-Out Structures (Smurfing)

Whereas circular structures are characterised by the return of capital, smurfing is defined by a fundamental *degree asymmetry*.

Deprez et al. [2025a] distinguish two canonical variants: in the *scatter-gather* topology, funds fan out from a source account via mule accounts and fan in at the destination; in the *gather-scatter* topology, funds are first collected at a central hub before being redistributed [Deprez et al., 2025a, p. 2]. A defining characteristic is the uni-directional flow of funds with minimal return flow [Deprez et al., 2025a, p. 6].

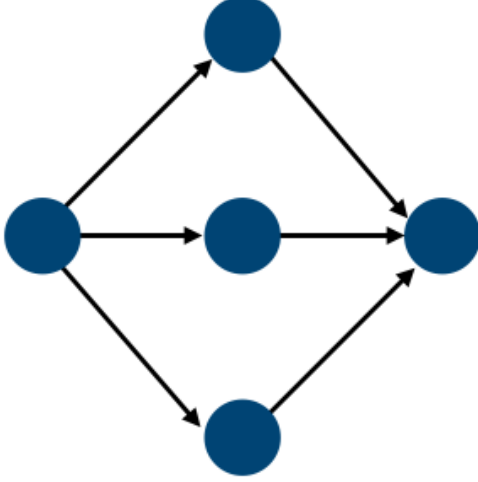
As illustrated in Figure 2.2, both topologies produce a characteristic degree asymmetry: high out-degree nodes (sources) and high in-degree nodes (destinations) are linked exclusively through intermediate mule nodes of low degree.

This structural signature distinguishes smurfing from the cyclic patterns of Section 2.3.1 and motivates the dedicated fan-in and fan-out degree features of Definition 3.2.6 (Section 3.2) and the detection features of Section 4.8.2.

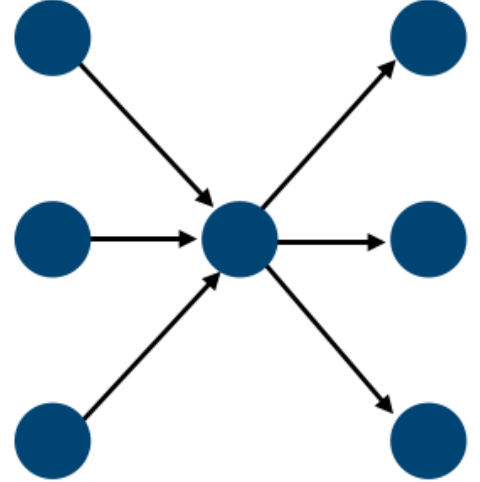
2.3.3 Layering Cascades

The layering phase routes funds through a series of complex financial transactions designed to confuse the audit trail, with the explicit aim of making the transaction trail challenging to follow [Association of Certified Fraud Examiners (ACFE), 2025a, Bonato et al., 2024, p. 3].

From a graph-theoretic perspective, layering cascades correspond to *bottleneck structures* in the transaction graph: subgraphs through which all fund flows must pass and that are connected to the rest of the network by relatively few edges. The spectral formalisation of this bottleneck property, via the algebraic connectivity of the symmetrised Laplacian and its relationship to isoperimetric constants, is developed in Sections 3.3.3 and 4.4.3.



(a) Scatter-gather: funds fan out from a single source and converge at a single destination.



(b) Gather-scatter: funds are first pooled at a central hub before being redistributed to multiple destinations.

Figure 2.2: The two canonical smurfing topologies. Based on [Deprez et al. \[2025a, Figure 1, p. 2\]](#).

2.3.4 Strongly Connected Components and Circulation

When multiple layering stages are coupled through feedback, *strongly connected components* (SCCs) emerge: subgraphs in which every node can reach every other node via directed paths (Definition 3.3.7).

[Granados and Vargas \[2022\]](#) formalise the underlying suspiciousness rules in terms of graph-structural properties: maximal cliques in the reduced network contain several intersecting cycles and thus satisfy all suspiciousness rules simultaneously [[Granados and Vargas, 2022](#), pp. 10–11]. Their empirical analysis of the Bahamas Leaks and Panama Papers networks further reveals negative degree assortativity [[Granados and Vargas, 2022](#), p. 4], indicating that high-degree nodes tend to connect predominantly to low-degree nodes, a structural pattern consistent with hub-and-spoke money-laundering behaviour.

Crucially, strongly connected components are an inherently directed notion and cannot, in general, be recovered from the kernel of the symmetrised Laplacian L^{sym} (Definition 3.2.13). Indeed, symmetrisation discards arc orientation, so that $\ker(L^{\text{sym}})$ reflects only the connected components of the undirected graph G_{sym} , equivalently the weakly connected components of the original transaction graph (Proposition 3.2.15). Consequently, strong connectivity is treated in this thesis purely as a combinatorial property of the directed transaction graph D (Section 5.3.3), while the spectral analysis

of connectivity and cyclic structure via L^{sym} is carried out on the symmetrised operator in Chapter 5, Section 5.4.1.

2.3.5 Temporal Bursts and Dynamic Graphs

Fontes et al. [2021] investigate motif structures in banking transaction graphs using a six-hour look-back window and show that the dominant fraud motifs consist of repeated transactions between the same client–merchant pairs within a single window, which appear as three-node cliques in the transaction graph [Fontes et al., 2021, p. 5]. This pattern supports the view that temporal concentration is a primary structural signature of fraudulent activity. It is, moreover, distinct from the path-based propagation patterns that emerge in the corresponding entity graph.

Weber et al. [2018] validate dynamic graph topologies on AMLSim, generating a synthetic transaction network with around one million nodes and nine million edges [Weber et al., 2018, pp. 4–5]. This scale illustrates that temporal organisation is not an optional refinement but an essential dimension of any realistic AML graph model. Abrupt changes in cyclic structure across successive time windows are therefore captured as a spectral feature in Section 4.8.1.

Summary

Table 2.1: Graph-theoretic patterns and their spectral and topological signals.

Section	Pattern	Mathematical signal	Formalised in
2.3.1	Cycles / rings	$\beta_1 \geq 1$, small $\lambda_1(L^{\text{sym}})$	Secs. 3.3.3 , 3.4.4
2.3.2	Fan-in / fan-out	$d_i^{\text{in}} \gg d_i^{\text{out}}$ (or vice versa), degree asymmetry (weighted volumes, Def. 3.2.6)	Def. 3.2.6 ; Sec. 4.8.2
2.3.3	Layering cascades	small $\lambda_1(L^{\text{sym}})$, Cheeger constant	Thm. 4.4.9
2.3.4	SCCs (directed) / circulation	SCCs: purely combinatorial in D ; circulation in G_{sym} : $\dim \ker(L^{\text{sym}}) > 1$	Sec. 5.3.3 ; Sec. 3.3.2
2.3.5	Temporal bursts	$\mathcal{F}_\beta(\tau) \geq 1$	Def. 4.8.1 , (4.45)

The operators L^{sym} , β_1 , λ_1 , $\ker(\tilde{L})$, and $\Delta_q^{K,L}$, together with the symmetrised graph G_{sym} are introduced in Chapter [3](#) and the temporal burst indicator $\mathcal{F}_\beta(\tau)$ in Section [4.8](#).

2.3.6 Bridging the Gap: From Directed Flows to Symmetrised Topological Shapes

A central challenge in mapping money-laundering typologies onto algebraic topology is the inherent directedness of financial transactions. As discussed in Section [2.2.1](#), placement, layering, and integration are directional processes in which capital moves sequentially from source to destination. Classical and persistent spectral theory for simplicial complexes, however, has largely been developed for undirected structures, where self-adjoint operators ensure real and stable spectra.

To bridge this conceptual gap, the present thesis adopts a deliberate abstraction step. Before constructing the clique complex $K(G_{\text{sym}})$ (Definition [3.4.4](#)), the directed transaction network is transformed into a symmetrised graph G_{sym} . This transformation deliberately discards the orientation of individual transactions and thereby changes the analytical focus from directed flows to undirected structural co-occurrences and topological backbones.

Directed layering cycles or multi-account smurfing hubs, for example, still appear as dense, tightly connected components or closed loops in G_{sym} . Although the precise directional flow is temporarily

abstracted away, the structural anomaly itself is retained and encoded in the combinatorial topology, for instance through non-vanishing Betti numbers β_k . The crucial temporal and volumetric dynamics of these directed flows are then recovered not via asymmetric operators, but through the multi-scale filtration process introduced in Chapter 4.

2.4 Existing Approaches and the Case for Spectral Methods

The development of AML detection methods can broadly be divided into three successive generations: rule-based systems, machine-learning approaches, and graph-based or spectral methods. In this section, each generation is reviewed in turn, with an emphasis on its structural limitations, and three core requirements are distilled that motivate the spectral-topological framework developed in Chapters 3 and 4.

2.4.1 Rule-Based Approaches

Rule-based TMS have long been the dominant AML technology. Their operating principle is straightforward: a predefined set of rules is evaluated for each incoming transaction, and an alert is generated whenever a threshold is violated [Bakhshinejad et al., 2024, p. 496]. The key operational drawback is an extremely high false-positive rate: the false alert rate (FAR) typically lies between 95 % and 98 % [Bakhshinejad et al., 2024, p. 497], a finding corroborated by Columbia School of International and Public Affairs (SIPA) [2018], who report values in the range of 90 %–95 % for industry-standard systems. This limitation is rooted in the architecture of rule-based approaches themselves: they rely on fixed scalar thresholds and are not designed to capture graph-topological structure such as cycles, hubs, or bipartite patterns.

2.4.2 Machine Learning and Graph-based Methods and Remaining Gaps

Graph-based methods are particularly well suited to AML, as transaction data are inherently network-structured [Weber et al., 2018, p. 3]. Even without machine learning, classical graph algorithms such as cycle detection, PageRank, degree-distribution analysis, and label propagation can reduce false positives by 20–30 % in systems where KYC information supports reliable entity identification [Weber et al., 2018, p. 3].

Granados and Vargas [2022] demonstrate empirically that topological analysis of large-scale financial networks can uncover suspicious clusters and describe how their members are separated, but also note that many laundering schemes rely on a small number of interactions within tightly knit groups that evade standard community-detection techniques [Granados and Vargas, 2022, pp. 17–18].

Graph neural networks go some way towards addressing this by exploiting network structure directly: Bakhshinejad et al. [2024] propose a graph convolutional network in which convolutions are interpreted as first-order approximations of local spectral filters on the graph [Bakhshinejad et al., 2024, p. 500]. Nonetheless, Deprez et al. [2025b], in a survey of 97 studies, conclude that most existing work still depends on expert-defined rules and manual feature engineering [Deprez et al., 2025b, p. 1], and that GNNs tend to struggle under extreme class imbalance and in hub-dominated networks [Deprez et al., 2025b, p. 29].

Important gaps therefore remain. In particular, most current approaches do not examine the stability of topological features across multiple filtration scales. Persistent Betti numbers are rarely employed in AML settings and the spectral signatures of directed graphs are largely unexplored in this application domain. Addressing these gaps is the central objective of Chapters 3 and 4. The persistent Laplacian $\Delta_q^{K,L}$ provides a scale-stable operator whose kernel recovers persistent Betti numbers along filtrations of transaction graphs, thereby combining spectral and topological information within a single self-adjoint operator.

2.4.3 Why Spectral and Topological Methods?

The limitations surveyed above converge to a common conclusion: effective AML detection requires methods that (i) sensitively capture higher-order structural patterns such as cycles and communities, (ii) characterise topological features in a stable way across multiple scales, and (iii) remain compatible with supervisory expectations on explainability.

Structural sensitivity. Spectral quantities such as the eigenvalues of the graph Laplacian $L = \mathcal{D} - W$ capture global network properties, including connectivity, community structure, and bipartiteness, that are inaccessible to local, threshold-based rules [von Luxburg, 2007, pp. 3–4]. Their formal development is deferred to Section 3.2.

Multi-scale robustness. Mémoli et al. [2022] introduce the persistent Laplacian $\Delta_q^{K,L}$, Wei and Wei [2024] further develop and apply it. Stability results for persistent eigenvalues show that controlled perturbations of the underlying filtration lead only to correspondingly bounded spectral changes [Mémoli et al., 2022, p. 22, Theorem 5.9].

Since transaction networks are inherently directed, Li and Zhang [2012] provide the necessary Laplacian theory for digraphs. This framework is recalled in Section 3.2 and integrated with the persistent-Laplacian construction in Chapter 4.

Interpretability and regulatory alignment. Spectral quantities such as $\lambda_1(L)$ and $\beta_q^{K,L}$ can be directly related to intuitive concepts such as connectivity, cycle density, and bottleneck strength, and can therefore be documented in a regulatory context, satisfying the explainability requirements recalled in Section 2.1 [European Parliament and Council \[2015\]](#), [Federal Republic of Germany \[2017\]](#). By contrast, the latent representations of deep neural networks are difficult to interpret under severe class imbalance and hub-dominated topologies, and they do not naturally expose topological invariants such as β_1 or λ_1 .

Chapter 3 develops the operator-theoretic foundation for all three properties: directed and symmetrised graph Laplacians, Hodge Laplacians on clique complexes, and the persistent Laplacian $\Delta_q^{K,L}$ as the central operator of Chapter 4.

The three requirements derived above, structural sensitivity, multi-scale robustness, and regulatory interpretability, translate directly into mathematical desiderata: a self-adjoint Laplacian whose kernel encodes cycle counts (β_1), whose smallest positive eigenvalue measures bottleneck strength (λ_1), and whose spectral perturbation is controlled by the interleaving distance between filtrations.

Chapter 3

Graph, Spectral, and Topological Foundations

The purpose of this chapter is to establish the mathematical foundations required for a spectral-topological analysis of transaction networks. Transaction data are modelled as directed, weighted multigraphs, reflecting the inherent asymmetry and heterogeneity of monetary flows. Building on this representation, the chapter develops a sequence of linear and topological operators that progressively enrich the structural information extracted from the network.

Section 3.1 introduces transaction graphs as directed multigraphs with weighted arcs and without loops, capturing both the structure and magnitude of financial interactions. Section 3.2 translates these discrete objects into matrix form via adjacency, degree, and Laplacian operators. While directed Laplacians preserve flow orientation, their lack of symmetry motivates the introduction of a symmetrised Laplacian, which is self-adjoint and therefore amenable to classical spectral theory.

Section 3.3 discusses connectivity, paths, and cycles, preparing the transition to higher-order structures. Section 3.4 introduces simplicial complexes and boundary operators, enabling the formulation of Hodge Laplacians in Section 3.5. These operators decompose flows into gradient, curl, and harmonic components, providing a multi-dimensional analogue of graph Laplacians.

Finally, Section 3.6 introduces persistent Laplacians along filtrations of simplicial complexes. These operators combine the stability of persistent homology with the analytical richness of spectral theory, yielding scale-dependent eigenvalues that quantify the persistence and strength of topological features. This framework prepares Chapter 4.

Throughout this chapter, all graphs are assumed to be finite. Unless stated otherwise, weights are assumed to be non-negative, and all vector spaces are considered over $\mathbb{R}$.

3.1 Graphs for Transaction Networks

The first step towards a spectral-topological analysis of transaction data is to specify an appropriate graph model.

Transaction networks are inherently asymmetric (sender $\rightarrow$ receiver), weighted (transaction amounts) and may contain multiple transactions and self-transfers. This motivates the following unified graph model as the basic mathematical object of the thesis.

Definition 3.1.1 (Transaction graph).

A *transaction graph* is defined as a finite directed weighted multigraph, i.e. a triple

$$D = (V, A, w),$$

where

- V is a non-empty finite set of vertices,
- $A \subseteq V \times V$ is a multiset of directed arcs, allowing multiple arcs between the same ordered pair but no loops of the form (v, v) ,
- $w : A \rightarrow \mathbb{R}_{\geq 0}$ is a non-negative weight function.

An arc $(u, v) \in A$ represents a directed transaction from account u to account v over a fixed observation window, with $w(u, v)$ denoting the associated transaction amount.

Remark 3.1.2. The above definition corresponds to a directed multigraph *without loops*: multiple arcs between the same ordered pair are permitted, but self-arcs of the form (v, v) are excluded from the outset. This suffices to accurately model transaction data, where multiple transfers between the same accounts are common, while self-transfers (an account sending funds to itself) are irrelevant to the detection of money-laundering cycles. Indeed, a loop (v, v) does *not* constitute a circular money flow between distinct accounts. The money-laundering cycles of interest such as the ring $A \rightarrow B \rightarrow C \rightarrow A$ discussed in Example 3.1.2, involve directed cycles through *at least two distinct* accounts. For all spectral and topological constructions in Sections 3.2–3.6, loops are therefore excluded. This convention is standard in spectral graph theory [Spielman, 2025, Naderi and Pankrashkin, 2024] and entails no loss of information relevant to the detection of cyclic fraud patterns, as those are captured via the first Betti number β_1 of the associated simplicial complex (see Section 3.4).

3.1.1 Directed graphs for transactions

Definition 3.1.3 (Graph).

A *graph* is an ordered pair $G = (V, E)$, where V is a non-empty and finite set of elements called

vertices, and E is a finite set of elements called *edges*. Each edge is a two-element subset $\{u, v\}$ with $u, v \in V$.

In an *undirected graph*, the pair set of E has no order: $\{u, v\} = \{v, u\}$ for $u, v \in V$ with $u \neq v$.

Such undirected graphs model symmetric relations. In contrast, monetary flows have a natural direction from sender to receiver, which is encoded by directed graphs.

Definition 3.1.4 (Directed graph, digraph).

A *directed graph* (or *digraph*) is an ordered pair $D = (V, A)$, where V is a non-empty finite set of vertices and $A \subseteq V \times V$ is a finite set of arcs.

An arc $(u, v) \in A$ is directed from the *tail* u to the *head* v .

Two vertices $u, v \in V$ are *adjacent* if $(u, v) \in A$ or $(v, u) \in A$.

Remark 3.1.5 (AML interpretation). In the AML setting, vertices represent accounts and arcs represent observed transactions:

$$V = \{\text{accounts}\}, \quad A = \{\text{observed transactions between accounts}\}.$$

An arc (u, v) corresponds to at least one transaction from account u to account v .

Definition 3.1.6 (In-degree and out-degree, semi-degrees).

Let $D = (V, A)$ be a digraph and $v \in V$. The *out-degree* of v is

$$d_D^+(v) := \#\{(v, u) \in A : u \in V\},$$

and the *in-degree* of v is

$$d_D^-(v) := \#\{(u, v) \in A : u \in V\}.$$

Remark 3.1.7 (Interpretation in transaction networks). In a transaction network, $d_D^+(v)$ counts the number of outgoing transactions from account v , while $d_D^-(v)$ counts incoming transactions.

Strong imbalances between in-degree and out-degree are often an indicator for collection or distribution accounts in money-laundering schemes, see Chapter 2.

3.1.2 Multi-edge and weighted models

Real transaction data may contain multiple transfers between the same pair of accounts and self-transfers. To capture this, multiple arcs and weights are allowed.

Definition 3.1.8 (Multigraph / Pseudograph).

A *multigraph* or *pseudograph* is a graph $G = (V, E)$ in which multiple edges between the same pair of vertices and loops (edges of the form $\{u, u\}$) are allowed.

Definition 3.1.9 (Directed Multigraph).

A *directed multigraph* is a digraph $D = (V, A)$ in which multiple arcs between the same ordered pair (u, v) are permitted.

Remark 3.1.10. Throughout this thesis, *transaction graphs* (Definition 3.1.1) are directed multigraphs allowing multiple arcs but *no* loops (v, v) . This differs from the general pseudograph convention [Bang-Jensen and Gutin, 2009, p. 18], in which loops are permitted; here they are excluded from the outset, since self-transfers are irrelevant to cyclic AML-pattern detection (Remark 3.1.2). All spectral and topological constructions in Sections 3.2–3.6 are therefore defined directly on $D = (V, A, w)$ without further loop-reduction.

In the following sections, a transaction graph containing the raw transactions is assumed. In this case, parallel arcs are sometimes combined into a single weighted arc.

Definition 3.1.11 (Weighted graph and weighted digraph).

A *weighted graph* is a triple $G = (V, E, w)$, where (V, E) is a graph and $w : E \rightarrow \mathbb{R}_{\geq 0}$ assigns a non-negative *weight* to each edge.

Similarly, a *weighted digraph* is a triple $D = (V, A, w)$ with $w : A \rightarrow \mathbb{R}_{\geq 0}$.

In the unweighted case, a weight of 1 is assumed.

For a transaction graph $D = (V, A, w)$, the weight $w(u, v)$ will typically denote the aggregated transaction volume from u to v over a fixed time window,

$$w(u, v) = \text{total transaction amount from account } u \text{ to account } v,$$

or, alternatively, the number of transactions or a risk-weighted amount. Parallel arcs corresponding to multiple transactions between the same accounts can either be stored explicitly or aggregated into a single weighted arc; both viewpoints lead to the same matrix representations considered in Section 3.2.

This representation retains both the structure (who transacts with whom) and the strength (how much is transferred) of connections and forms the basis for the spectral and topological analysis in the subsequent sections.

Example: Transaction network To illustrate the above notions, consider a example transaction network with four accounts $V = \{A, B, C, D\}$ observed over a fixed time window. Accounts A, B, C form a directed transfer ring with identical transaction volumes, while account D only engages in a few low-volume transfers. Table 3.1 summarises the aggregated transaction amounts between the accounts.

In this example, the high-volume ring $A \rightarrow B \rightarrow C \rightarrow A$ represents a stylised cyclic flow pattern, as it is often observed in simplified money-laundering schemes [Association of Certified Fraud Examiners (ACFE), 2025a, Granados and Vargas, 2022], whereas the edges $A \rightarrow D$ and $D \rightarrow C$ model comparatively harmless transfers. The full transaction graph is thus given by the directed weighted

Table 3.1: Example transaction graph with four accounts and aggregated transaction volumes.

Sender	Receiver	Weight $w_{u,v}$	Interpretation
A	B	5000	part of ring $A \rightarrow B \rightarrow C \rightarrow A$
B	C	5000	part of ring $A \rightarrow B \rightarrow C \rightarrow A$
C	A	5000	part of ring $A \rightarrow B \rightarrow C \rightarrow A$
A	D	100	isolated low-volume transfer
D	C	200	isolated low-volume transfer

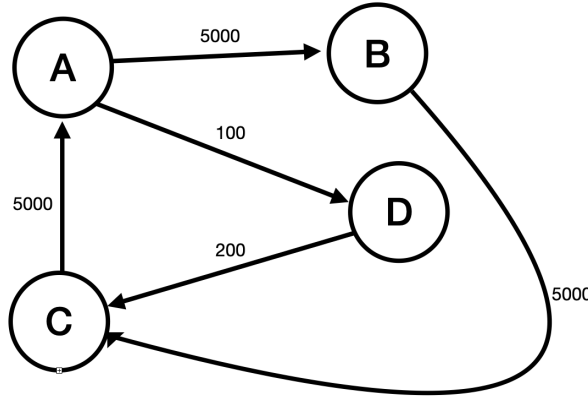


Figure 3.1: Graph representation of the example transaction network $D = (V, A, w)$ with vertices $V = \{A, B, C, D\}$ and weighted directed edges as in Table 3.1.

multigraph $D = (V, A, w)$ with the vertex set $V = \{A, B, C, D\}$ and arc weights as in Table 3.1.

While such a tabular representation is convenient for small networks, it quickly becomes unintuitive once the number of accounts and transactions increases: individual entries are hard to relate to each other and global structures such as rings, hubs, or cascades are no longer visually apparent. This motivates representing the same data as a graph, in which vertices correspond to accounts and directed, weighted edges encode the observed transactions. Figure 3.1 illustrates the toy network from Table 3.1 in this graphical form.

In this thesis, the detailed modelling of real-world transaction data as directed, weighted graphs is discussed in Chapter 4. The present chapter focuses solely on introducing and motivating the mathematical foundations of graph-based, spectral, and topological methods that will later be applied to such transaction networks.

3.1.3 Local structure: subgraphs and strongly connected regions

Many money-laundering patterns are localized: they involve small groups of accounts that form tightly connected rings, layered paths or cascades; see Chapter 2. Such patterns are conveniently described via subgraphs of the transaction graph.

Definition 3.1.12 (Subgraph and induced subgraph).

Let $G = (V, E)$ be a (directed or undirected) graph.

- A *subgraph* of G is an ordered pair $G' = (V', E')$ with $V' \subset V$ and $E' \subset E$.
- For $U \subset V$, the *subgraph of G induced by U* is defined as $G_U := (U, E_U)$ with

$$E_U := \{\{u, u'\} \in E : u, u' \in U\}$$

in the undirected case, and analogously $E_U := \{(u, u') \in E : u, u' \in U\}$ in the directed case.

Remark 3.1.13 (Use in AML subgraph analysis). In practice, the clique complexes and associated Hodge and persistent Laplacians introduced in Chapters 3 and 4 are computed not on the full transaction graph, but on induced subgraphs G_U corresponding to subsets of accounts flagged as suspicious by rule-based systems or preliminary machine-learning models. This restriction to candidate subgraphs controls the combinatorial growth of higher-order cliques in dense regions (e.g. around exchanges or payment service providers) and makes the persistent Laplacian layer computationally feasible. For the topological constructions in Sections 3.4–3.6, only the underlying undirected connectivity of these subgraphs is relevant, obtained either by forgetting orientations or via the symmetrisation procedure introduced in Section 3.2.

3.2 Classical Graph Laplacians and Spectra

This section is primarily based on von Luxburg [2007], Spielman [2025], and Naderi and Pankrashkin [2024]; directed Laplacians follow Prayitno et al. [2019] and Li and Zhang [2012].

One of the most powerful tools in graph theory is the translation of combinatorial structures into linear algebra. By associating graphs with matrices, access to spectral theory is gained: eigenvalues, eigenvectors, and quadratic forms become representatives of connectivity, flow, and clustering. As Spielman [2025, p. 3] puts it, a matrix can be understood as a *linear operator* that maps a vector $\mathbf{v}$ to $M\mathbf{v}$, or as a *quadratic form* that maps $\mathbf{v}$ to the scalar $\mathbf{v}^\top M\mathbf{v}$.

In this section, adjacency and degree matrices, as well as the corresponding Laplacian matrices, are introduced. These matrices form the central link between the combinatorial structure of transaction graphs and the spectral methods that are later generalised to Hodge and persistent Laplace operators. In the undirected setting, graph Laplacians are real symmetric and positive semidefinite, with well-known connections between their spectra and notions of connectivity, clustering, and flow.

From a modelling perspective, transaction networks are inherently directed and may contain multiple edges; loops are excluded from all spectral and topological constructions as explained in Remark 3.1.2. Most existing theory for Hodge Laplacians and persistent Laplacians has been developed for symmetric Laplacians on undirected weighted graphs. Therefore a two-layer perspective is adopted in this section: on the one hand, directed Laplacians based on in- and out-degrees preserve information about flow orientation and local imbalances of funds, but are typically non-symmetric and thus fall outside the standard spectral framework; on the other hand, a symmetrised Laplacian is derived from the directed transaction graph, which fits into the classical symmetric setting while retaining undirected weighted connectivity information. This symmetrised operator serves as the basis for the topological and persistent Laplace operators considered in later sections.

Remark 3.2.1 (Notation). Throughout this section, the following notation is used:

- D denotes a directed graph (digraph) as in Definition 3.1.1;
- $\mathcal{A}$ denotes the (unweighted) adjacency matrix of an undirected graph; W denotes the weighted adjacency matrix, both for undirected and directed graphs;
- $\mathcal{D}$ denotes a diagonal degree matrix (with superscripts $\mathcal{D}^{\text{in}}$, $\mathcal{D}^{\text{out}}$, $\mathcal{D}^{\text{sym}}$ in the directed and symmetrised settings);
- L denotes the combinatorial Laplacian of an undirected graph;
- $\mathcal{L}$ denotes the normalised Laplacian;
- L_{sym} denotes the symmetrised Laplacian associated with a directed graph.
- $d_D^-(v)$, $d_D^+(v)$ denote the *unweighted* in- and out-degrees (arc counts) of Definition 3.1.6; d_i^{in} , d_i^{out} denote the corresponding *weighted* degrees (arc-weight sums) of Definition 3.2.6. The two coincide when $w \equiv 1$.

Throughout this work, we restrict ourselves to finite graphs. Hence, all associated matrices define bounded linear operators on *finite-dimensional* spaces.

3.2.1 Adjacency and Degree matrices

We begin with the standard definitions for undirected weighted graphs.

Definition 3.2.2 ((Weighted) adjacency matrix).

Let $G = (V, E)$ be a simple undirected graph with $n = |V|$ vertices. Fix an ordering $V = \{v_1, \dots, v_n\}$. The *adjacency matrix* of G is the $n \times n$ matrix $\mathcal{A} = (\mathcal{A}_{ij})$ defined by

$$\mathcal{A}_{ij} := \begin{cases} 1 & \text{if } \{v_i, v_j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

In the *weighted* case, $\mathcal{A}$ is replaced by a matrix $W = (w_{ij})$ with non-negative entries w_{ij} that encode edge weights [von Luxburg, 2007, p. 2].

In the finite-dimensional setting considered here, the adjacency matrix of an undirected graph defines a bounded self-adjoint operator on $\mathbb{R}^n$.

Definition 3.2.3 (Degree matrix).

Let $G = (V, E)$ be an undirected graph with $n = |V|$ vertices. The *degree matrix* of G is the diagonal $n \times n$ matrix $\mathcal{D} = (\mathcal{D}_{uv})_{u,v \in V}$ defined by

$$\mathcal{D}_{uv} := \begin{cases} d(u) & \text{if } u = v, \\ 0 & \text{otherwise,} \end{cases}$$

where $d(u)$ denotes the (possibly weighted) degree of vertex u .

As the primary goal of this thesis is to model financial transactions, we must transition from undirected structures to directed ones. Standard directed Laplacians are typically non-symmetric and possess complex spectra, which prevents a direct application of classical spectral clustering and robust multi-scale homological tools. To ensure algebraic and spectral stability along the topological filtrations developed in Chapter 4, a two-layer operator perspective is adopted.

In the first layer, the directed multi-edge data is consolidated into un-symmetrised matrix representations to compute local flow imbalances and semi-degrees. In the second layer, these representations are mapped onto a self-adjoint, symmetrised operator. This procedural step effectively yields a real, non-negative spectrum that is robust against small volume perturbations, while directional flow information is retained separately in the original directed graph and its directed operators. The simplicial complex constructed from the symmetrised graph is used exclusively for the undirected topological and persistent-spectral analysis. Consequently, we now establish the corresponding directed definitions for a transaction graph $D = (V, A, w)$ from Definition 3.1.1.

Definition 3.2.4 (Adjacency matrix of a directed multigraph).

Let $D = (V, A)$ be a directed multigraph with vertex set $V = \{v_1, \dots, v_n\}$, where loops are excluded. The *weighted adjacency matrix* of D is the matrix $W = (w_{ij})_{1 \leq i, j \leq n}$ defined by

$$w_{ij} := \sum_{e \in A(v_i, v_j)} w(e)$$

where $A(v_i, v_j)$ denotes the set of all arcs from v_i to v_j . If D is unweighted, we set $w(e) = 1$ for all $e \in A$, reducing w_{ij} to the arc multiplicity $\mu(v_i, v_j)$.

In particular, for *transaction graphs* the entry w_{ij} can be taken as the sum of the weights of all arcs from v_i to v_j .

Remark 3.2.5. Note that throughout this thesis, the symbol W is used uniformly for the weighted adjacency matrix, both in the undirected setting (Definition 3.2.2) and in the directed setting (Definition 3.2.4). The unweighted adjacency matrix of an undirected graph is denoted $\mathcal{A}$.

In the transaction-network setting, the weighted adjacency matrix W may correlate with either the number of transactions between accounts in a given time window (if arcs are counted) or the total transferred amounts (if arc weights are summed). The degree matrix collects the total outgoing or incident volume per account.

To define Laplacians in the directed setting, we also require in- and out-degree matrices consistent with W .

Definition 3.2.6 (Degree matrices of a (multi)digraph).

Let $D = (V, A, w)$ be a finite directed (multi)graph with vertex set $V = \{v_1, \dots, v_n\}$, arc multiset A (loops excluded), and a nonnegative weight function $w : A \rightarrow \mathbb{R}_{\geq 0}$. Define the weighted adjacency matrix $W = (w_{ij})_{1 \leq i, j \leq n}$ as in Definition 3.2.4. For each vertex $v_i \in V$ define the *in-degree* and *out-degree* as

$$d_i^{\text{in}} := \sum_{j=1}^n w_{ji}, \quad d_i^{\text{out}} := \sum_{j=1}^n w_{ij},$$

for directed graphs.

These quantities extend the unweighted arc-counts of Definition 3.1.6 to the weighted setting: whereas $d_D^-(v)$ and $d_D^+(v)$ count arcs, d_i^{in} and d_i^{out} aggregate arc *weights* (transaction volumes). The two coincide when $w \equiv 1$.

The corresponding diagonal *in-/out-degree matrices* $\mathcal{D}^{\text{in}}$ and $\mathcal{D}^{\text{out}}$ are defined by

$$\mathcal{D}_{ii}^{\text{in}} = d_i^{\text{in}}, \quad \mathcal{D}_{ii}^{\text{out}} = d_i^{\text{out}}, \quad \mathcal{D}_{ij}^{\text{in}} = \mathcal{D}_{ij}^{\text{out}} = 0 \quad (i \neq j).$$

Remark 3.2.7 (Flow interpretation). In a transaction graph, d_i^{out} collects the total outgoing transaction volume (or count) from account v_i , and d_i^{in} the incoming volume, over the chosen time window. These quantities will later be used to quantify local flow imbalances that are characteristic for money mules, collection hubs or cash-out accounts in laundering schemes.

3.2.2 Graph Laplacians

Spectral methods on graphs are built around Laplacian matrices, whose eigenvalues and eigenvectors encode connectivity and flow structure in a network [von Luxburg, 2007, p. 3].

Definition 3.2.8 (Combinatorial Laplacian).

Let $G = (V, E)$ be an undirected weighted graph with degree matrix $\mathcal{D}$ and weighted adjacency matrix W . The *(combinatorial) graph Laplacian* of G is the $n \times n$ matrix

$$L := \mathcal{D} - W.$$

Definition 3.2.9 (Normalized Laplacian).

Let G be as above and assume that all vertex degrees are positive. The *normalized Laplacian* of G is

$$\mathcal{L} := I - \mathcal{D}^{-1/2} W \mathcal{D}^{-1/2},$$

where I denotes the identity matrix.

For both L and $\mathcal{L}$, edges with large weight contribute strongly to the corresponding quadratic forms $x^\top Lx$ and $x^\top \mathcal{L}x$, penalizing differences between the values of x on adjacent vertices.

This makes Laplacians natural operators for measuring the smoothness of signals on graphs: a cochain f has small Rayleigh quotient $f^\top Lf$ precisely when it varies little across densely connected regions of the graph [von Luxburg, 2007, p. 28]. From a financial-crime perspective, these operators provide the bridge between the combinatorial structure of transaction graphs and quantitative indicators such as cluster separation and connectivity of suspicious sub-networks [von Luxburg, 2007]; the additional indicator of cyclic money flows is captured by the 1-Hodge Laplacian L_1 and its kernel, as detailed in Sections 3.5 and 3.6.

Motivated by transaction networks adapt these constructions to the directed setting. A directed Laplacian preserves information about flow orientation but is typically non-symmetric, whereas Hodge and persistent Laplacians require symmetric, positive semidefinite operators. This leads to a two-layer approach: we keep directed Laplacians for flow-based analysis and introduce a *symmetrised Laplacian* for the topological and persistent-spectral layers.

Definition 3.2.10 (Directed Laplacians).

The *in-degree Laplacian* and *out-degree Laplacian* of a directed (multi)graph $D = (V, A)$ (loops excluded) are defined as

$$L^{\text{in}}(D) := \mathcal{D}^{\text{in}} - W, \quad L^{\text{out}}(D) := \mathcal{D}^{\text{out}} - W.$$

Remark 3.2.11. By the loop-exclusion convention of Remark 3.1.2, any self-transfer arc (v, v) contributes 0 to the weighted adjacency matrix W of D , and hence does not affect the directed Laplacians L^{in} and L^{out} .

Example: Directed Laplacians for the example transaction graph We now compute the directed Laplacians for the transaction graph introduced in Example 3.1.2. Fix the vertex ordering $V = (A, B, C, D)$ and let $W = (w_{ij})_{i,j=1}^4$ denote the weighted adjacency matrix of the corresponding directed multigraph D . According to Definition 3.2.4, the entries w_{ij} collect the aggregated transaction volumes from account i to account j over the observation window. For the example network in Table 3.1 this yields

$$W = \begin{pmatrix} 0 & 5000 & 0 & 100 \\ 0 & 0 & 5000 & 0 \\ 5000 & 0 & 0 & 0 \\ 0 & 0 & 200 & 0 \end{pmatrix},$$

where, for instance, $w_{12} = 5000$ corresponds to the aggregated transfers from A to B , and $w_{34} = 0$ reflects the absence of a direct transaction from C to D .

Following Definition 3.2.6, the out-degree and in-degree of a vertex v_i are obtained as $d_{\text{out}}(v_i) = \sum_j w_{ij}$ and $d_{\text{in}}(v_i) = \sum_j w_{ji}$, respectively. For the present example one obtains

$$d_{\text{out}}(A) = 5100, \quad d_{\text{out}}(B) = 5000, \quad d_{\text{out}}(C) = 5000, \quad d_{\text{out}}(D) = 200,$$

and

$$d_{\text{in}}(A) = 5000, \quad d_{\text{in}}(B) = 5000, \quad d_{\text{in}}(C) = 5200, \quad d_{\text{in}}(D) = 100.$$

Hence the diagonal out-degree and in-degree matrices $\mathcal{D}^{\text{out}}$ and $\mathcal{D}^{\text{in}}$ take the form

$$\mathcal{D}^{\text{out}} = \begin{pmatrix} 5100 & 0 & 0 & 0 \\ 0 & 5000 & 0 & 0 \\ 0 & 0 & 5000 & 0 \\ 0 & 0 & 0 & 200 \end{pmatrix}, \quad \mathcal{D}^{\text{in}} = \begin{pmatrix} 5000 & 0 & 0 & 0 \\ 0 & 5000 & 0 & 0 \\ 0 & 0 & 5200 & 0 \\ 0 & 0 & 0 & 100 \end{pmatrix}.$$

By Definition 3.2.10, the out-degree and in-degree Laplacians are given by $L^{\text{out}} = \mathcal{D}^{\text{out}} - W$ and $L^{\text{in}} = \mathcal{D}^{\text{in}} - W$, i.e.

$$L^{\text{out}} = \begin{pmatrix} 5100 & -5000 & 0 & -100 \\ 0 & 5000 & -5000 & 0 \\ -5000 & 0 & 5000 & 0 \\ 0 & 0 & -200 & 200 \end{pmatrix},$$

$$L^{\text{in}} = \begin{pmatrix} 5000 & -5000 & 0 & -100 \\ 0 & 5000 & -5000 & 0 \\ -5000 & 0 & 5200 & 0 \\ 0 & 0 & -200 & 100 \end{pmatrix}.$$

Remark 3.2.12 (AML interpretation of directed Laplacians). The matrices L^{out} and L^{in} reflect local imbalances between outgoing and incoming transaction volumes at each account. The diagonal entries of $\mathcal{D}^{\text{out}}$ encode total outflow per account, while the off-diagonal entries of L^{out} reflect redistribution along directed edges. These operators serve as the directed-graph layer of the two-layer approach introduced in Definition 3.2.13; the topological layer is built on L^{sym} .

In general, the matrices $L^{\text{in}}(D)$ and $L^{\text{out}}(D)$ are not symmetric, since the adjacency matrix W of a directed graph is asymmetric whenever the arc set is not invariant under arc reversal. Consequently, their spectra may contain complex eigenvalues and many of the classical spectral properties of undirected Laplacians (real, nonnegative spectrum, variational characterisations, Cheeger-type inequalities, etc.) do not carry over directly [Chung, 2005, pp. 381–383][Li and Zhang, 2012, pp. 1–7].

Definition 3.2.13 (symmetrised graph and Laplacian).

Let $D = (V, A, w)$ be a finite directed weighted multigraph and let $n = |V|$ denote the number of vertices. Let $W \in \mathbb{R}^{n \times n}$ be the corresponding weighted adjacency matrix of D . The *symmetrised weighted adjacency matrix* is defined by

$$W^{\text{sym}} := \frac{1}{2}(W + W^\top).$$

The *symmetrised graph* is the undirected weighted graph

$$G_{\text{sym}} := (V, E_{\text{sym}}, W^{\text{sym}}),$$

where E_{sym} contains all unordered pairs $\{i, j\}$ such that $W_{ij}^{\text{sym}} > 0$. The associated degree matrix $\mathcal{D}^{\text{sym}}$ is defined by

$$\mathcal{D}_{ii}^{\text{sym}} := \sum_{j=1}^n W_{ij}^{\text{sym}}, \quad \mathcal{D}_{ij}^{\text{sym}} := 0 \quad (i \neq j),$$

and the *symmetrised Laplacian* is

$$L_{\text{sym}} := \mathcal{D}^{\text{sym}} - W^{\text{sym}}.$$

Remark 3.2.14. The matrix L_{sym} is the combinatorial Laplacian of the undirected weighted graph G_{sym} . In particular, it is real symmetric and positive semi-definite, as follows directly from the quadratic-form characterisation (3.1) applied to G_{sym} , and its spectral properties can be analysed using standard results from spectral graph theory [von Luxburg, 2007, pp. 4–6][Spielman, 2025, pp. 5–6].

The Filtration as a Proxy for Directed Dynamics. The potential information loss incurred by this graph symmetrisation is systematically mitigated by the multi-scale filtration architecture developed in detail in Section 3.6.2. As formalised in Definition 3.6.1, a simplicial filtration allows us to

examine a family of growing complexes $\mathcal{K} = \{K_t\}_{t \in T}$ indexed by a continuous or discrete parameter $t \in \mathbb{R}$. In a financial-crime context, this sequence is generated by thresholding the symmetrised edge-weight matrix W^{sym} along transaction volumes or frequencies.

Proposition 3.2.15 (Basic properties of the symmetrised Laplacian). *[von Luxburg, 2007, pp.4–6] Let $D = (V, A, w)$ be a finite directed weighted pseudograph and let L_{sym} be the symmetrised Laplacian associated with D .*

Then:

1. L_{sym} is a real symmetric matrix.
2. L_{sym} is positive semi-definite.
3. The multiplicity of the eigenvalue 0 equals the number of connected components of G_{sym} .

The proofs for those properties are given in von Luxburg [2007] [von Luxburg, 2007, pp.5–6].

3.2.3 Basic spectral properties

In this subsection we recall the key spectral properties of combinatorial and normalised Laplacians on undirected weighted graphs. By the discussion above, the same statements apply to the symmetrised Laplacian L_{sym} associated with a transaction graph D .

Theorem 3.2.16 (Symmetry and positive semi-definiteness). *[von Luxburg, 2007, Props. 1, 3, pp. 4 ff.] Let $G = (V, E)$ be an undirected weighted graph with weighted adjacency matrix $W = (w_{ij})_{1 \leq i, j \leq n}$. Let $\mathcal{D} = \text{diag}(d(v_1), \dots, d(v_n))$ be the diagonal degree matrix, where $d(v_i) := \sum_{j=1}^n w_{ij}$ denotes the weighted degree of vertex v_i (assumed strictly positive for all $i \in \{1, \dots, n\}$). Define the combinatorial Laplacian as $L := \mathcal{D} - W$ and the normalised Laplacian as $\mathcal{L} := I - \mathcal{D}^{-1/2} W \mathcal{D}^{-1/2}$. Then L and $\mathcal{L}$ are real symmetric matrices, and for all $x \in \mathbb{R}^n$,*

$$x^\top L x = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (x_i - x_j)^2 \geq 0, \quad (3.1)$$

$$x^\top \mathcal{L} x = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left(\frac{x_i}{\sqrt{d(v_i)}} - \frac{x_j}{\sqrt{d(v_j)}} \right)^2 \geq 0. \quad (3.2)$$

In particular, all eigenvalues of L and $\mathcal{L}$ are real and non-negative.

The proof for L is given in von Luxburg [2007, p. 4] and for $\mathcal{L}$ in von Luxburg [2007, pp. 5 f.]. Applied to L_{sym} , this shows that the symmetrised transaction Laplacian admits real, nonnegative eigenvalues and an orthogonal eigenbasis. These properties will later be mirrored in the Hodge Laplacians on

simplicial complexes built from G_{sym} , whose kernels and spectra encode higher-order structures such as cycles and cavities in transaction networks.

Theorem 3.2.17 (Zero eigenvalues and connected components). *[von Luxburg, 2007, Proposition 2, Proposition 4, pp. 4 ff.] Let $G = (V, E)$ be an undirected weighted graph with combinatorial Laplacian $L := \mathcal{D} - W$ and normalised Laplacian $\mathcal{L} := I - \mathcal{D}^{-1/2}W\mathcal{D}^{-1/2}$. Then:*

1. *The multiplicity of the eigenvalue 0 of L equals the number of connected components of G .*
2. *The same holds for $\mathcal{L}$: the multiplicity of the eigenvalue 0 of $\mathcal{L}$ equals the number of connected components of G .*

The proof for L is given in von Luxburg [2007, pp. 4 f.]; for $\mathcal{L}$ the proof is analogous using Theorem 3.2.16, under the assumption that all vertex degrees of G are positive (i.e. G has no isolated vertices).

For the symmetrised transaction graph G_{sym} , the multiplicity of the zero eigenvalue of L_{sym} thus counts the number of undirected connected components after symmetrization. In the AML setting, this allows us to distinguish disconnected clusters of accounts that do not share any transaction paths (even indirectly), which is essential when localizing suspicious subgraphs and interpreting persistent Laplacians on filtrations of G_{sym} .

The smallest eigenvalue of both L and $\mathcal{L}$ is always 0, and its eigenspace is spanned by indicator vectors of the connected components. This provides a first link between spectral information and global connectivity.

Theorem 3.2.18 (Spectral theorem for symmetric matrices). *Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric matrix. Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of M , and all eigenvalues of M are real.*

Remark 3.2.19. Various formulations and proofs of the spectral theorem can be found in standard university lecture notes and textbooks on linear algebra; see e.g. Naderi and Pankrashkin [2024, p. 28].

Applied to L and $\mathcal{L}$, the spectral theorem guarantees that these matrices admit an *orthogonal eigenvalue decomposition*. This property will later be mirrored in the Hodge Laplacians in Section 3.5, where the spectral decomposition may correlate with topological information such as the number of cycles.

Example: symmetrised Laplacian for the toy transaction graph We now compute the symmetrised Laplacian L_{sym} associated with the transaction graph from Example 3.1.2. Let W denote its weighted adjacency matrix as in Example 3.2.2. Following Definition 3.2.13, the symmetrised weighted adjacency matrix is

$$W^{\text{sym}} = \frac{1}{2}(W + W^{\top}).$$

For the present example one computes

$$W^{\top} = \begin{pmatrix} 0 & 0 & 5000 & 0 \\ 5000 & 0 & 0 & 0 \\ 0 & 5000 & 0 & 200 \\ 100 & 0 & 0 & 0 \end{pmatrix}, \quad W + W^{\top} = \begin{pmatrix} 0 & 5000 & 5000 & 100 \\ 5000 & 0 & 5000 & 0 \\ 5000 & 5000 & 0 & 200 \\ 100 & 0 & 200 & 0 \end{pmatrix},$$

and thus

$$W^{\text{sym}} = \begin{pmatrix} 0 & 2500 & 2500 & 50 \\ 2500 & 0 & 2500 & 0 \\ 2500 & 2500 & 0 & 100 \\ 50 & 0 & 100 & 0 \end{pmatrix}.$$

The corresponding symmetrised degree matrix $\mathcal{D}^{\text{sym}}$ has diagonal entries $d_{\text{sym}}(v_i) = \sum_j (W^{\text{sym}})_{ij}$, i.e.

$$d_{\text{sym}}(A) = 5050, \quad d_{\text{sym}}(B) = 5000, \quad d_{\text{sym}}(C) = 5100, \quad d_{\text{sym}}(D) = 150,$$

and consequently

$$\mathcal{D}^{\text{sym}} = \begin{pmatrix} 5050 & 0 & 0 & 0 \\ 0 & 5000 & 0 & 0 \\ 0 & 0 & 5100 & 0 \\ 0 & 0 & 0 & 150 \end{pmatrix}.$$

The symmetrised combinatorial Laplacian L_{sym} is then

$$L_{\text{sym}} = \mathcal{D}^{\text{sym}} - W^{\text{sym}} = \begin{pmatrix} 5050 & -2500 & -2500 & -50 \\ -2500 & 5000 & -2500 & 0 \\ -2500 & -2500 & 5100 & -100 \\ -50 & 0 & -100 & 150 \end{pmatrix}.$$

By construction, L_{sym} is a real symmetric, positive semi-definite matrix, so that all its eigenvalues are real and non-negative and an orthonormal eigenbasis exists, in accordance with Theorems 3.2.16 and 3.2.18. In the symmetrised view, the high-volume ring A – B – C appears as a strongly weighted triangle, while D is only weakly attached to this core, reflecting its peripheral role in the toy transaction network.

3.3 Paths, Connectivity, and Cycles

The combinatorial material in this section follows primarily [Naderi and Pankrashkin \[2024\]](#) and [Bang-Jensen and Gutin \[2009\]](#); spectral connectivity results are taken from [Spielman \[2025\]](#).

Money-laundering schemes are often built from characteristic structures such as long transfer chains, tightly connected rings of accounts, and densely interlinked subnetworks. These patterns are naturally expressed using the language of walks, paths, cycles, and connectivity concepts for graphs. In transaction networks, funds flow along directed paths from senders to recipients and often return to their source within multi-layered structures or closely interconnected rings; such configurations underlie typical money laundering schemes and fraud rings, in which sums of money are routed through chains of intermediary accounts and ultimately concentrated again; see Chapter 2.

Building on the transaction graph model from Section 3.1 and the Laplace operators introduced in Section 3.2, the notions of walks, paths, connectivity, and cycles are formalised in this section for both the original directed transaction graph and its symmetrised counterpart. Directed paths and cycles encode the possible routes along which funds can move and return, while connectivity and components describe how the universe of accounts decomposes into subnetworks. Spectral indicators such as the multiplicity of the zero eigenvalue and the algebraic connectivity link these combinatorial properties to the spectrum of the (symmetrised) Laplacian. The same ideas reappear later in the language of Hodge-Laplace operators and persistent Laplace operators, where paths and cycles correspond to higher-order harmonic structures.

For the combinatorial definitions, both the directed transaction graph $D = (V, A, w)$ and its symmetrised, undirected counterpart G_{sym} , associated with L_{sym} , are used. The former reflects the actual direction of cash flows, while the latter provides the symmetric framework required for the spectral and topological tools developed in Sections 3.4–3.6.

Remark 3.3.1 (Spectral viewpoint in the symmetrised setting). In the following, all spectral statements about Laplacians (such as the characterisation of connected components via the multiplicity of the eigenvalue 0 or the interpretation of the first positive eigenvalue as algebraic connectivity) are understood for the *symmetrised transaction graph* G_{sym} and its Laplacian L_{sym} , which is real symmetric and positive semidefinite.

They do not apply to the non-symmetric in-/out-degree Laplacians L^{in} and L^{out} of the original directed multigraph D , which are used solely to encode oriented flow information.

3.3.1 Walks, Paths, and Cycles

On graphs, walks and paths provide the basic language to describe movement and cyclic structures.

Definition 3.3.2 (Undirected walks, paths, cycles).

Let $G = (V, E)$ be an undirected graph and let $x, y \in V$, $k \in \mathbb{N}$.

- A *walk* of length k from x to y is a sequence $(x_0, x_1, \dots, x_k)$ of vertices such that $x_0 = x$, $x_k = y$, and $\{x_{j-1}, x_j\} \in E$ for all $j = 1, \dots, k$.
- The walk is *closed* if $x = y$.
- If all vertices in the sequence are pairwise distinct, the walk is called a *path*.
- A closed walk $(x_0, \dots, x_k)$ with $x_0 = x_k$ and $x_0, \dots, x_{k-1}$ pairwise distinct is called a *cycle*.

When reasoning about undirected connectivity it is sufficient to consider simple paths, since every walk between two vertices contains a path between them.

Proposition 3.3.3. [*Naderi and Pankrashkin, 2024*, p. 9, Proposition 1.1.14]

Let $G = (V, E)$ be an undirected graph and let $u, v \in V$, $u \neq v$. If there exists a walk from u to v , then there also exists a path from u to v .

The above notions transfer to *financial graphs* (see Definition 3.1.1) by requiring that all edges are traversed in their orientation.

Definition 3.3.4 (Directed walks, trails, paths, cycles).

Let $D = (V, A)$ be a directed multigraph.

- A *directed walk* in D is a finite sequence

$$W = x_1 a_1 x_2 a_2 \dots a_{k-1} x_k$$

of vertices $x_i \in V$ and arcs $a_j \in A$ such that, for every $i = 1, \dots, k-1$, the tail of a_i is x_i and the head of a_i is x_{i+1} .

- The *length* of W is the number of its arcs, i.e. $k-1$. The walk is called *closed* if $x_1 = x_k$ and *open* otherwise. When the arcs are clear from the context, we write $W = x_1 x_2 \dots x_k$.
- A *trail* is a walk in which all arcs are distinct.
- A walk (or trail) is called a *path* if all its vertices are distinct.
- A closed walk $W = x_1 \dots x_k$ with $k \geq 3$, $x_1 = x_k$ and $x_1, \dots, x_{k-1}$ pairwise distinct is called a *cycle*.

In the AML context, directed cycles in the transaction graph D model typical fraud rings: money leaves a source account, passes through a series of intermediary accounts, and returns to the source or to a tightly connected group of accounts [Granados and Vargas, 2022, pp. 8–9, Rules 3 and 5]. In Section 3.5 these cycles will be represented as elements of the kernel of the Hodge 1-Laplacian on edges, and in Section 3.6 persistent Laplacians will quantify the stability of such cyclic structures across filtration scales.

Example: Directed cycle in the example transaction graph In the transaction graph from Example 3.1.2, the vertices A, B, C and the arcs $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow A$ form a directed cycle of length three. More precisely, the sequence

$$A \rightarrow B \rightarrow C \rightarrow A$$

is a closed directed walk in the sense of Definition 3.3.4, and since all intermediate vertices are distinct, it is a directed cycle. In the AML context, this cycle models a simplified fraud ring in which funds can circulate indefinitely among the three accounts via repeated layering transactions [Association of Certified Fraud Examiners (ACFE), 2025a, Layering], while account D remains only weakly connected to this core structure via the low-volume edges $A \rightarrow D$ and $D \rightarrow C$, consistent with structuring patterns designed to stay below reporting thresholds [Association of Certified Fraud Examiners (ACFE), 2025a, Placement].

3.3.2 Connectivity and Components

Connectivity formalises the intuitive notion that a graph consists of one piece or several disjoint parts.

Definition 3.3.5 (Connected graph and distance).

Let $G = (V, E)$ be an undirected graph. The graph G is called *connected* if for any two distinct vertices $u, v \in V$ there exists a path from u to v . The *distance* $d(u, v)$ between u and v is defined as the length of a shortest path between u and v , and $d(u, u) := 0$.

Definition 3.3.6 (Connected components).

A *connected component* of an undirected graph $G = (V, E)$ is a maximal connected subgraph of G . If G has components $G_1, \dots, G_p$ with vertex sets $V_1, \dots, V_p$ forming a partition of V , this is often denoted by

$$G = G_1 \sqcup \dots \sqcup G_p.$$

After reordering the vertices so that those belonging to the same component are grouped together, the adjacency and Laplacian matrices of G acquire a block-diagonal form, and each block acts on one component.

For the symmetrised transaction graph G_{sym} associated with L_{sym} , this decomposition reflects a partition of the account universe into undirected components between which there is no path, even if directed transactions exist internally in each component.

In *directed* graphs, connectivity must take edge directions into account. The *directed distance* $\text{dist}(u, v)$ from u to v is defined as the length of a shortest directed path from u to v , with $\text{dist}(u, u) := 0$; if no directed path from u to v exists, $\text{dist}(u, v) := \infty$.

Definition 3.3.7 (Strong connectivity and SCCs).

A directed graph $D = (V, A)$ is called *strongly connected* or *strong* if for every pair of vertices $u, v \in V$ there exists a directed path from u to v :

$$\exists m \in \mathbb{N}, \exists x_0, \dots, x_m \in V: x_0 = u, x_m = v, \text{ and } (x_{i-1}, x_i) \in A \text{ for all } 1 \leq i \leq m.$$

A *strongly connected component (SCC)* of D is a maximal strongly connected subgraph of D .

Remark 3.3.8 (SCCs in transaction networks). In transaction networks, SCCs represent groups of accounts between which funds can circulate along directed paths. Classic fraud rings correspond to SCCs that contain non-trivial directed cycles and unusually dense internal transfer patterns while remaining weakly connected to the rest of the graph [Granados and Vargas, 2022, pp. 8–9]. In Chapter 4 these notions are used to design subgraphs on which Hodge and persistent Laplacians are computed.

3.3.3 Spectral Indicators of Connectivity

The graph Laplacian provides a direct spectral handle on connectivity. For an undirected weighted graph, the multiplicity of the zero eigenvalue of the Laplacian equals the number of connected components, and the smallest positive eigenvalue (the Fiedler value, Definition 3.3.10) quantifies how well these components are tied together [von Luxburg, 2007, pp. 4–6]. In the present setting, these statements apply in particular to the symmetrised Laplacian L_{sym} of a transaction graph.

Proposition 3.3.9 (Smallest eigenvalue of L). [Naderi and Pankrashkin, 2024, p. 36] Let $G = (V, E)$ be a graph with connected components $G_j = (V_j, E_j)$, $j = 1, \dots, N$. Then 0 is the smallest eigenvalue of the Laplacian L and

$$\dim \ker L = N, \quad \ker L = \text{span}\{\mathbf{1}_{V_j} : j = 1, \dots, N\}.$$

In particular, G is connected $\Leftrightarrow 0$ is a simple eigenvalue of L .

The proof is found in [Naderi and Pankrashkin, 2024, p. 36].

Applied to the symmetrised transaction Laplacian L_{sym} , this means that its kernel may correlate with the coarse decomposition of the account universe into undirected connected components [von Luxburg,

2007, pp. 5–6, Prop. 4][Lim, 2020, p. 17].

In spectral AML applications, this allows one to detect disconnected clusters of accounts from the spectrum alone [Granados and Vargas, 2022, pp. 4–5], and to restrict attention to components that contain suspicious activity when constructing filtrations for persistent Laplacians.

For the Laplacian of a graph, the first non-zero eigenvalue characterises how tightly the graph is connected [Lim, 2020, p. 17].

Definition 3.3.10 (Algebraic connectivity / Fiedler value).

Let $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ denote the eigenvalues of the Laplacian L (or of $\mathcal{L}$) of a graph $G = (V, E)$. The *algebraic connectivity* or *Fiedler value* of G is defined as the smallest positive eigenvalue λ_1 .

Lemma 3.3.11. [Spielman, 2025, Lemma 3.1.1, p. 30] *Let $G = (V, E)$ be an undirected weighted graph with Laplacian L or $\mathcal{L}$. Let $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ denote the eigenvalues of the Laplacian. Then*

$$\lambda_1 > 0 \iff G \text{ is connected (Definition 3.3.5).}$$

The proof can be found in [Spielman, 2025, pp. 30–31].

For the normalised Laplacian, small values of λ_1 indicate that the graph can be partitioned into two large vertex sets separated by relatively few edges, while large values of λ_1 imply the absence of such sparse cuts. Cheeger-type inequalities make this intuition precise by relating λ_1 to isoperimetric ratios of cuts in the graph; see Spielman [2025, pp. 171 ff.] for details.

In Section 3.5, analogous statements will be formulated for the Hodge 1-Laplacian L_1 acting on edges: its kernel provides a proxy for cycles in the sense of 1-homology, and its smallest positive eigenvalue measures how robust one-dimensional structures (such as transaction rings and cascades) are. In Section 3.6, persistent Laplacians will inherit these interpretations across filtration scales, allowing us to distinguish transient from structurally persistent patterns in transaction networks.

3.4 Simplicial Complexes and Boundary Operators

The material in this section follows primarily Mémoli et al. [2022] and Lim [2020]. Clique complexes are treated in Lim [2020, p. 6].

Graphs capture pairwise relationships between accounts, but many money-laundering patterns involve coordinated behaviour within small groups of accounts, such as tightly knit rings or clusters of intermediaries. To model such higher-order interactions, the setting is extended from graphs to

simplicial complexes, in which vertices, edges, and triangles encode multi-way relationships.

In this section, simplicial complexes are constructed from the symmetrised transaction graph G_{sym} , which is associated with the Laplace operator L_{sym} introduced in Section 3.2. This topological framework forms the basis for Hodge-type Laplace operators in Section 3.5 and later for persistent Laplace operators on transaction networks in Section 3.6.

3.4.1 Simplices and Simplicial Complexes

Throughout this section, the discussion is limited to dimensions $k \leq 2$, which are sufficient for the later use of 0- and 1-Hodge and persistent 1-Hodge Laplacians on transaction graphs.

Definition 3.4.1 (k -simplices).

Let V be a finite set. A non-empty subset $\sigma \subseteq V$ with $|\sigma| = k + 1$ is called a k -*simplex*. Thus, 0-simplices are single vertices, 1-simplices are unordered edges, and 2-simplices are triangles.

Definition 3.4.2 (Simplicial complex).

A non-empty family K of finite subsets of V is an (*abstract*) *simplicial complex* if it is closed under inclusion: whenever $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$. The elements of K are called *simplices*, the collection of all k -simplices is denoted K_k .

In particular, if $\{u, v, w\}$ is a 2-simplex in K , then $\{u, v\}$, $\{u, w\}$, $\{v, w\}$ and the three vertices themselves must also belong to K .

A convenient way to build a simplicial complex from an undirected graph is via cliques.

Definition 3.4.3 (Cliques in a graph).

Let $G = (V, E)$ be an undirected graph with finite vertex set $V = \{1, \dots, n\}$ and edge set $E \subseteq \binom{V}{2}$. For $k \geq 1$, the set of k -*cliques* is

$$\text{Cl}_k(G) := \{\{i_1, \dots, i_k\} \subseteq V : \{i_p, i_q\} \in E \text{ for all } 1 \leq p < q \leq k\},$$

i.e. $\text{Cl}_k(G)$ consists of all k -element vertex sets that span complete subgraphs in G .

Definition 3.4.4 (Clique complex).

Let $G = (V, E)$ be a graph with clique sets $\text{Cl}_k(G)$ and clique number $\omega(G)$, the size of a largest clique in G . The *clique complex* of G is the simplicial complex

$$K(G) := \bigcup_{k=1}^{\omega(G)} \text{Cl}_k(G),$$

where the collection of k -simplices of $K(G)$ (as defined in Definition 3.4.2) is given by $K_k := \text{Cl}_{k+1}(G)$.

In $K(G)$, vertices of G become 0-simplices, edges become 1-simplices, and each triangle of mutually connected vertices becomes a 2-simplex [Lim, 2020, p. 6].

For transaction networks, we will apply this construction to the symmetrised graph G_{sym} , the clique complex $K(G_{\text{sym}})$ thus lifts the (symmetrised) transaction structure to a higher-dimensional object in which tightly coupled groups of accounts become explicit topological entities, suitable for homological and spectral analysis.

Example: Clique complex of the example graph Let G_{sym} denote the symmetrised transaction graph of the Example 3.2.3, obtained from the weighted adjacency matrix W^{sym} . Since the accounts A, B, C are mutually connected in G_{sym} , they form a 3-clique and thus give rise to a 2-simplex $\{A, B, C\}$ in the clique complex $K(G_{\text{sym}})$.

In particular, $K(G_{\text{sym}})$ contains the 0-simplices A, B, C, D , the 1-simplices corresponding to the edges of G_{sym} , and the single 2-simplex $\{A, B, C\}$ spanning the high-volume transaction triangle. From a topological perspective, the fraud ring $A-B-C$ becomes an explicit 2-dimensional simplex in $K(G_{\text{sym}})$, on which homology and Hodge Laplacians can be defined.

3.4.2 Chain Groups

To perform algebraic constructions on a simplicial complex, such as the definition of boundary operators or Laplace operators, we work with formal linear combinations of simplices organized into chain groups.

Definition 3.4.5 (Oriented simplices and chain groups).

Let K be a simplicial complex on a finite ordered vertex set V equipped with a total ordering $<$. An *oriented q -simplex* is a simplex $\sigma = \{v_0, \dots, v_q\} \in K$ with vertices ordered as $v_0 < v_1 < \dots < v_q$, written $[v_0, \dots, v_q]$.

The q -th chain group $C_q^K := C_q(K, \mathbb{R})$ is the real vector space freely generated by all oriented q -simplices, subject to the relation

$$[v_{\pi(0)}, \dots, v_{\pi(q)}] = \text{sgn}(\pi) [v_0, \dots, v_q]$$

for every permutation π of $\{0, \dots, q\}$.

In the graph case, C_0^K consists of formal linear combinations of vertices, C_1^K of oriented edges, and C_2^K of oriented triangles. These spaces support the boundary operators used to define homology and later Hodge Laplacians on the clique complex $K(G_{\text{sym}})$.

The total ordering $<$ on V imposes a formal reference orientation on the simplices of $K(G_{\text{sym}})$, decoupled from the physical direction of money flows in the directed transaction graph $D = (V, A, w)$. The directed transaction data are encoded as follows.

Definition 3.4.6 (Transaction 1-chain). Let $D = (V, A, w)$ be a directed weighted transaction graph and $K(G_{\text{sym}})$ its clique complex with total vertex ordering $<$. The *transaction 1-chain* associated with D is

$$\omega_D := \sum_{(u,v) \in A, u < v} w(u,v) [u, v] - \sum_{(u,v) \in A, u > v} w(u,v) [v, u] \in C_1(K(G_{\text{sym}})).$$

When the Hodge 1-Laplacian L_1 is applied to ω_D in Chapter 4, its kernel captures cyclic money-flow patterns and its nonzero spectrum encodes geometric flow information, while net-flow asymmetries at vertices are handled by the directed Laplacians of Section 3.2. For the applications in this thesis, we work with the chain groups $C_0(K(G_{\text{sym}}))$, $C_1(K(G_{\text{sym}}))$, and $C_2(K(G_{\text{sym}}))$.

3.4.3 Boundary Operators

The topological structure of a simplicial complex is encoded in boundary operators on chains, which connect adjacent dimensions and satisfy the fundamental “boundary of a boundary” identity.

Definition 3.4.7 (Boundary operators).

Let K be a simplicial complex with chain groups C_q^K . For $q \geq 1$, the q -th *boundary operator* $\partial_q^K : C_q^K \rightarrow C_{q-1}^K$ is defined on an oriented q -simplex $[v_0, \dots, v_q] \in K$ by

$$\partial_q^K[v_0, \dots, v_q] := \sum_{i=0}^q (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_q],$$

where $\hat{v}_i$ indicates omission of the i -th vertex. The definition is extended linearly to all of C_q^K .

In particular, the boundary of an oriented edge is the alternating sum of its endpoints, and the boundary of an oriented triangle is the alternating sum of its oriented edges. For all q it holds that

$$\partial_{q-1}^K \circ \partial_q^K = 0,$$

reflecting the fact that the boundary of a boundary is empty.

Example: Path Let G be the path on three vertices $v_0 - v_1 - v_2$ and $K = K(G)$ its clique complex. Then $C_0(K)$ is spanned by v_0, v_1, v_2 and $C_1(K)$ by $e_{01} = [v_0, v_1]$ and $e_{12} = [v_1, v_2]$. With the ordering $v_0 < v_1 < v_2$,

$$\partial_1(e_{01}) = v_1 - v_0, \quad \partial_1(e_{12}) = v_2 - v_1.$$

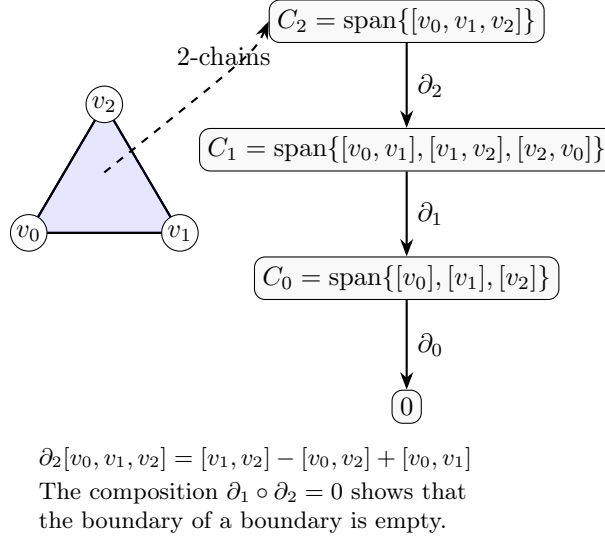


Figure 3.2: Chain complex of a 2-simplex and its boundary operators.

Example: Triangle Let G be a triangle on vertices v_0, v_1, v_2 and $K = K(G)$. Then $C_2(K)$ is spanned by $t = [v_0, v_1, v_2]$ and $C_1(K)$ by $e_{01} = [v_0, v_1]$, $e_{02} = [v_0, v_2]$, $e_{12} = [v_1, v_2]$. With $v_0 < v_1 < v_2$,

$$\partial_2(t) = e_{12} - e_{02} + e_{01}, \quad \partial_1(e_{ij}) = v_j - v_i.$$

These examples already illustrate how 1-cycles can appear as 1-chains with zero boundary (for instance, suitable linear combinations of edges around a triangle). This perspective will reappear in the interpretation of the kernel of the 1-Hodge Laplacian and, later, of the persistent 1-Laplacian on filtrations of $K(G_{\text{sym}})$.

3.4.4 Homology and Betti Numbers

Homology groups quantify the presence of holes of different dimensions in a simplicial complex. In the context of transaction networks, the corresponding Betti numbers measure, in particular, how many connected components and independent cyclic flow patterns are present.

Definition 3.4.8 (Homology groups and Betti numbers).

Let K be a simplicial complex with boundary operators ∂_q^K . Since $\partial_q^K \circ \partial_{q+1}^K = 0$ [Lim, 2020, p. 13, Eq. (4.4)], the image $\text{im}(\partial_{q+1}^K)$ is a subspace of $\ker(\partial_q^K)$. For $q \geq 0$, the q -th homology group of K is

$$H_q(K) := \ker(\partial_q^K) / \text{im}(\partial_{q+1}^K).$$

Its dimension

$$\beta_q^K := \dim H_q(K)$$

is called the q -th Betti number of K .

β_q^K counts the number of independent q -dimensional holes in K [Lim, 2020, pp. 10–12, 16–17]. In transaction graphs, β_0 measures disconnected account clusters in G_{sym} and $\beta_1 \geq 1$ signals at least one independent cyclic flow pattern—a topological signature of layering and fraud rings Section 2.2) [Granados and Vargas, 2022, pp. 8–9]. The spectral characterisation $\beta_k = \dim \ker(L_k)$ is established in Theorem 3.5.10 below.

The following standard consequence of the rank–nullity theorem for linear maps between finite-dimensional vector spaces will be used throughout Chapters 4 and 5.

Corollary 3.4.9 (Rank–nullity for boundary operators). *Let K be a finite simplicial complex and $\partial_q^K : C_q^K \rightarrow C_{q-1}^K$ its q -th boundary operator. Then the rank–nullity relation*

$$\dim C_q^K = \text{rank}(\partial_q^K) + \text{nullity}(\partial_q^K)$$

holds. In particular, the q -th Betti number satisfies

$$\beta_q^K = \text{nullity}(\partial_q^K) - \text{rank}(\text{im } \partial_{q+1}^K),$$

where $\beta_q^K = \dim H_q(K)$ is defined in Definition 3.4.8.

In the next section, these homological invariants will be linked directly to the spectra of Hodge Laplacians, providing the key bridge from topology to spectral features that will eventually be exploited in persistent Laplacian-based AML indicators.

3.5 The Hodge Laplacian

This section follows primarily Lim [2020], combinatorial Laplacians on chains are treated in Mémoli et al. [2022, pp. 4–5].

The material in this section establishes the non-persistent case: the Hodge Laplacian Δ_q^K on a single simplicial complex K , whose kernel recovers the classical Betti number β_q^K (Theorem 4.2.10). Section 3.6 then generalises this construction to a simplicial pair $X \subset Y$, yielding the persistent Laplacian $\Delta_q^{X,Y}$ whose kernel recovers the persistent Betti number $\beta_q^{X,Y}$ (Theorem 3.6.8).

Throughout this section, chain and cochain spaces are identified via the standard ℓ^2 inner product; under this identification $\Delta_q^K = L_q$, as made precise in the following remark.

Remark 3.5.1 (Chain–cochain identification). Throughout this thesis, the canonical identification

$$C_q(K) \cong C^q(K), \quad \sigma_i \mapsto \sigma_i^*,$$

between chain and cochain spaces via the ℓ^2 inner product (Remark 3.5.2) is used without further comment. Under this identification, boundary and coboundary operators are related by $\delta_q = \partial_{q+1}^*$ (Definition 3.5.3), and the combinatorial Laplacian on chains Δ_q^K coincides with the Hodge q -Laplacian on cochains L_q . In particular, $H_q(K) \cong H^q(K)$ as real vector spaces under this identification, so that Theorems 3.5.10 and 3.5.11 are equivalent formulations of the same spectral fact.

The classical graph Laplacian from Section 3.2 acts on functions on vertices and describes connectivity and flow structure at the level of edges. In the simplicial setting, an analogous construction exists for k -cochains, leading to Hodge Laplacians that capture not only vertex-based diffusion but also higher-order patterns such as cycles and surface-like structures.

Throughout this section, the clique complex $K = K(G_{\text{sym}})$ of the symmetrised transaction graph introduced in Sections 3.1–3.4 is considered. Hodge Laplacians on these clique complexes are defined, their fundamental properties are recalled, and their relationship to homology via Betti numbers is explained. The 0-Hodge Laplacian recovers the usual graph Laplacian, whereas the 1-Hodge Laplacian acts on edge-based flows and admits a discrete Helmholtz decomposition into gradient, curl, and harmonic components [Lim, 2020, pp. 10–11].

In money laundering, the layering stage directs illicit funds through multiple parties via complex transaction flows to conceal their origin [Kurshan and Shen, 2020, p. 7]. The Hodge decomposition provides a natural mathematical framework to separate potential-driven flows from cyclic transfer patterns, and to identify persistent harmonic flow structures that may be indicative of such layering activity.

3.5.1 Chain and cochain spaces, inner products, and adjoints

Let K be a finite simplicial complex. For $q \geq 0$, the q -th chain group $C_q(K)$ is the real vector space spanned by oriented q -simplices of K , see Section 3.4. The corresponding cochain space is the dual space

$$C^q(K) := \text{Hom}(C_q(K), \mathbb{R}).$$

Here, $\text{Hom}(V, W)$ denotes the real vector space of linear maps from a finite-dimensional real vector space V to a finite-dimensional real vector space W . In particular, $C^q(K)$ is the space of linear functionals on the chain group $C_q(K)$. In finite dimension, $C^q(K)$ is again a finite-dimensional real vector space.

Remark 3.5.2. Following Mémoli et al. [2022, p. 5] and Lim [2020, p. 7], all chain and cochain spaces are equipped with inner products defined with respect to the canonical basis of oriented simplices

and the corresponding dual basis, respectively. Concretely, if $\{\sigma_1, \dots, \sigma_m\}$ denotes the set of oriented q -simplices, the standard ℓ^2 -inner product of [Lim \[2020, p. 7, Eq. \(3.1\)\]](#) declares $\{\sigma_1, \dots, \sigma_m\}$ to be an orthonormal basis of $C_q(K)$. Adjoints of linear operators between chain or cochain spaces are always understood with respect to these inner products.

For $q \geq 1$, the boundary operator $\partial_q : C_q(K) \rightarrow C_{q-1}(K)$ is defined as in Definition 3.4.7 in Section 3.4 and extended linearly [[Mémoli et al., 2022, pp. 4–5](#)]. The coboundary operators on cochains are then given by

$$\delta_q : C^q(K) \rightarrow C^{q+1}(K), \quad \delta_q := \partial_{q+1}^*,$$

where ∂_{q+1}^* denotes the adjoint of ∂_{q+1} with respect to the inner products specified in Remark 3.5.2 [[Mémoli et al., 2022, p. 5](#)]. This definition is consistent with the explicit formula for higher-order coboundary operators given in [Lim \[2020, p. 13, Eq. \(4.2\)\]](#). By convention, $\partial_0 := 0$ and $\delta_{-1} := 0$ [[Mémoli et al., 2022, p. 5](#)].

Definition 3.5.3 (Coboundary operator).

For $q \geq 0$, the q -th coboundary operator is defined as the adjoint of the boundary operator,

$$\delta_q : C^q(K) \rightarrow C^{q+1}(K), \quad \delta_q := \partial_{q+1}^*,$$

where ∂_{q+1}^* denotes the adjoint of ∂_{q+1} with respect to the inner products of Remark 3.5.2. By convention, $\partial_0 := 0$ and $\delta_{-1} := 0$.

This definition is consistent with the explicit formula for higher-order coboundary operators given in [Lim \[2020, p. 13, Eq. \(4.2\)\]](#).

Definition 3.5.4 (Adjoint coboundary operator). Let K be a finite simplicial complex with cochain spaces $C^k(K)$ equipped with the inner products of Remark 3.5.2. For $k \geq 0$, the *adjoint coboundary operator* is the map

$$\delta_k^* : C^{k+1}(K) \rightarrow C^k(K),$$

defined as the adjoint of $\delta_k : C^k(K) \rightarrow C^{k+1}(K)$ with respect to these inner products, i.e. the unique operator satisfying

$$\langle \delta_k f, g \rangle_{C^{k+1}(K)} = \langle f, \delta_k^* g \rangle_{C^k(K)} \quad \text{for all } f \in C^k(K), g \in C^{k+1}(K).$$

Since $C^k(K)$ and $C^{k+1}(K)$ are finite-dimensional inner product spaces, δ_k^* is uniquely determined and equals the matrix transpose of δ_k in the orthonormal bases of Remark 3.5.2. By convention, $\delta_{-1}^* := 0$.

3.5.2 Definition of the Hodge k -Laplacian

On the chain level, combinatorial Laplacians are built from boundary operators and their adjoints.

Definition 3.5.5 (Combinatorial Laplacian on chains).

Let K be a simplicial complex with chain groups C_q^K and boundary operators ∂_q^K . For $q \geq 0$, the q -th combinatorial Laplacian on chains is the operator

$$\Delta_q^K : C_q^K \rightarrow C_q^K, \quad \Delta_q^K := \underbrace{\partial_{q+1}^K (\partial_{q+1}^K)^*}_{\Delta_{q,+}^K} + \underbrace{(\partial_q^K)^* \partial_q^K}_{\Delta_{q,-}^K}.$$

For $q = 0$ and unit weights, Δ_0^K reduces to the classical combinatorial graph Laplacian on the 1-skeleton of K [Mémoli et al., 2022, p. 5]. In this sense, the combinatorial Laplacians extend the classical Laplacian to all dimensions in a unified framework.

3.5.3 Hodge k -Laplacians on cochains

On the cochain level, an analogous construction is obtained by using the coboundary operators δ_q introduced in Definition 3.5.3 in Section 3.5.1.

Definition 3.5.6 (Hodge k -Laplacian on cochains).

Let K be a simplicial complex with cochain spaces $C^k(K)$ and coboundary operators $\delta_k : C^k(K) \rightarrow C^{k+1}(K)$. For $k \geq 0$, the Hodge k -Laplacian is the self-adjoint operator

$$L_k : C^k(K) \rightarrow C^k(K), \quad L_k := \delta_{k-1} \delta_{k-1}^* + \delta_k^* \delta_k,$$

where by convention $\delta_{-1} := 0$. A k -cochain $f \in C^k(K)$ satisfying $L_k f = 0$ is called a *harmonic k -cochain*.

Self-adjointness and positive semidefiniteness of L_k follow directly from the definition as a sum of two positive semidefinite operators [Lim, 2020, p. 4, p. 14]. For $k = 0$, L_0 coincides with the graph Laplacian acting on vertex-based signals when K is the clique complex of a graph [Lim, 2020, p. 8]. For $k = 1$, L_1 acts on edge-based cochains and gives rise to the discrete Hodge decomposition of edge flows [Lim, 2020, pp. 10–11].

In this thesis, the main object of interest will be the 1-Hodge Laplacian L_1 on the clique complexes $K(G_{\text{sym}})$ of (symmetrised) transaction graphs, and later its persistent counterparts along time-based filtrations.

3.5.4 Harmonic Cochains and Hodge Decomposition

The Hodge Laplacians have a precise relationship with cohomology and admit an orthogonal decomposition of cochain spaces into exact, coexact, and harmonic parts.

Theorem 3.5.7 (Hodge decomposition).

[Lim, 2020, p. 14, Sec. 4.3, p. 30 (A.2)] *Let K be a finite simplicial complex and $k \geq 0$. The k -th cohomology group*

$$H^k(K) := \ker(\delta_k) / \text{im}(\delta_{k-1})$$

is isomorphic to the space of harmonic k -cochains,

$$H^k(K) \cong \ker(\delta_k) \cap \ker(\delta_{k-1}^*) = \ker(L_k),$$

and the cochain space $C^k(K)$ admits the orthogonal decomposition

$$C^k(K) = \text{im}(\delta_{k-1}) \oplus^\perp \ker(L_k) \oplus^\perp \text{im}(\delta_k^*),$$

where $\oplus^\perp$ denotes an orthogonal direct sum with respect to the standard inner product on $C^k(K)$.

The decomposition follows from an application of the abstract linear-algebraic Hodge theorem to the cochain complex: since $\delta_k \circ \delta_{k-1} = 0$ by [Lim, 2020, p. 13 (4.4)], one may set $A = \delta_k$ and $B = \delta_{k-1}$ in Lim [2020, p. 3–4, Sec. 2.3], from which the isomorphism and the orthogonal decomposition follow by Lim [2020, p. 30, Thm. A.2].

Elements of $\text{im}(\delta_{k-1})$ are referred to as *exact k -cochains*, elements of $\text{im}(\delta_k^*)$ as *coexact k -cochains*, and harmonic cochains form canonical representatives of cohomology classes. The decomposition shows that every cochain can be written uniquely as the sum of an exact, a coexact, and a harmonic part. For $k = 1$, this specialises to a discrete Helmholtz decomposition of edge flows.

Theorem 3.5.8 (Helmholtz decomposition for edge flows). [Lim, 2020, pp. 10–11] *Let G be a graph with at least one triangle and let $K(G)$ be its clique complex. Then the space of 1-cochains decomposes orthogonally as*

$$C^1(K(G)) = \text{im}(\text{grad}) \oplus \ker(L_1) \oplus \text{im}(\text{curl}^*).$$

Moreover,

$$\ker(L_1) = \ker(\text{curl}) \cap \ker(\text{div}).$$

In words, any edge-based flow can be decomposed into a gradient part (driven by a potential on vertices), a curl part (driven by triangular flux), and a harmonic part (cyclic flows not bounded by any triangle) [Lim, 2020, pp. 10–11, Table 2]. In the AML context, the harmonic component of the

1-Hodge Laplacian reflects precisely the independent cyclic money flows that are not explained by local potential differences or higher-order surface structures, providing a spectral signature of fraud rings [Lim, 2020, p. 17].

Example: Harmonic edge flow on the example ring Consider again the transaction graph on $\{A, B, C, D\}$ and the clique complex $K(G_{\text{sym}})$. Define an edge-based cochain $f \in C^1(K(G_{\text{sym}}))$ that assigns the same constant value $c > 0$ to each oriented edge along the four-cycle $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ and zero to all other edges.

Geometrically, f represents a purely cyclic flow circulating along the ring $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ without any source or sink at any vertex. Since no triangle in $K(G_{\text{sym}})$ covers all four edges simultaneously, the curl of f vanishes; and since every vertex has equal algebraic in-flow and out-flow with respect to this oriented cycle, the divergence of f vanishes as well. Hence $f \in \ker(\text{curl}) \cap \ker(\text{div}) = \ker(L_1)$ [Lim, 2020, pp. 11–12], illustrating how a circular money-flow pattern appears as a harmonic component in the Hodge decomposition of edge flows.

Remark 3.5.9. The cycle $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ is a directed cycle in the original transaction graph D . After symmetrisation, the four edges form a 1-cycle in $K(G_{\text{sym}})$. This cycle is harmonic precisely because it is *not* the boundary of any 2-chain in $K(G_{\text{sym}})$: no triangle fills in all four edges, so the cycle persists topologically as a generator of $H_1(K(G_{\text{sym}}))$ [Lim, 2020, pp. 11–12].

By contrast, the three-cycle $A \rightarrow B \rightarrow C \rightarrow A$ would *not* be harmonic, since $\{A, B, C\}$ is a 2-simplex in $K(G_{\text{sym}})$ (see Example 3.4.1) and the cycle is therefore the boundary of that triangle. The self-transfer (loop) case (v, v) , excluded by Remark 3.1.2, would contribute no 1-cycle and is therefore irrelevant for this analysis.

3.5.5 Betti Numbers and Spectra

This subsection follows Lim [2020, pp. 10–12, 16–17] and Mémoli et al. [2022, pp. 4–5].

The kernels of Hodge Laplacians encode homological information: their dimensions coincide with Betti numbers and therefore quantify the number of holes of different dimensions.

For clique complexes of graphs, the Betti numbers of Definition 3.4.8 admit a spectral characterisation via Hodge Laplacians.

Theorem 3.5.10 (Betti numbers via Hodge Laplacians). [Lim, 2020, pp. 10–12, 16–17]

Let G be a graph with Hodge k -Laplacians L_k on the clique complex $K(G)$. Then

$$\beta_k(K(G)) = \dim H^k(K(G)) = \dim \ker(L_k).$$

In particular, $\beta_0(G) = \dim \ker(L_0)$ equals the number of connected components of G , and $\beta_1(G) = \dim \ker(L_1)$ counts the number of independent one-dimensional cycles.

Proof. By Theorem 3.5.7, the Hodge isomorphism gives $H^k(K(G)) \cong \ker(L_k)$, so $\dim H^k(K(G)) = \dim \ker(L_k)$. Since $\beta_k = \dim H^k(K(G))$ by Definition 3.4.8, the first equality follows. The case $k = 0$ recovers the classical fact that $\dim \ker(L_0)$ equals the number of connected components of G , as constant functions on each component form a basis of $\ker(L_0)$ [Lim, 2020, p. 10]. For $k = 1$, the equality $\beta_1 = \dim \ker(L_1)$ counts independent cycles not bounded by any triangle, i.e. elements of $\ker(\text{curl}) \cap \ker(\text{div})$ [Lim, 2020, p. 12]. □

Thus, computing the spectrum of L_k simultaneously yields both geometric information via nonzero eigenvalues and eigenvectors and topological information via the multiplicity of the eigenvalue 0 [Lim, 2020, p. 17].

These relationships extend to combinatorial Laplacians on chains.

Theorem 3.5.11 (Nullity of combinatorial Laplacians). [*Mémoli et al., 2022, p. 5, Thm. 2.2*] Let K be a simplicial complex and Δ_q^K its q -th combinatorial Laplacian on chains. Then, for every $q \geq 0$,

$$\beta_q^K = \text{nullity}(\Delta_q^K).$$

Proof. The proof is made in Mémoli et al. [2022] [Mémoli et al., 2022, p. 28]. □

This result extends the familiar relationship between the kernel of the graph Laplacian and connectivity to all dimensions.

In the next section, this Hodge-theoretic viewpoint is combined with filtrations of simplicial complexes to define *persistent* Laplacians, whose kernels recover persistent homology and whose spectra provide additional information along the filtration. For transaction networks, this will allow us to track which topological features of money flows (such as rings or clusters) persist across scales and which appear only as transient noise.

3.6 Persistent Laplacians

This section introduces persistent Laplacians as a preview of Chapter 4. Definitions follow Mémoli et al. [2022] and Wei and Wei [2024]; proofs and algorithms are deferred to Chapter 4.

Classical homology and Hodge theory describe the topology of a single complex at a fixed scale. In transaction networks relevant structures may appear or disappear as thresholds in time, volume, or risk are varied, and the main interest lies in patterns that persist across a range of scales rather than in transient noise.

Classical persistent homology summarises how topological features such as connected components and cycles appear and disappear along a filtration, but it only records the dimensions of homology groups [Wei and Wei, 2024, pp. 3–4]. Persistent Laplacians refine this picture: their kernels recover persistent homology, while their non-zero eigenvalues encode additional geometric structure along the filtration [Wei and Wei, 2024, pp. 3–4, 10–11]. The full theory, algorithms, and application to transaction networks are developed in Chapter 4; the present section introduces the key definitions and motivates their use in the AML context.

Building on the Hodge-theoretic framework of Section 3.5, we now extend the construction to a pair of simplicial complexes $K \subseteq L$. The key idea is to replace the single boundary operator ∂_q^K by a restricted operator $\partial_q^{K,L}$ that “sees” the ambient complex L , while acting on the chain group $C_q(K)$ of the smaller complex.

3.6.1 Motivation: Limitations of Persistent Homology

Persistent homology describes how topological features, such as connected components, cycles, and voids, appear and disappear as a simplicial complex evolves along a filtration

$$K^0 \subseteq K^1 \subseteq \dots \subseteq K^N.$$

For each pair of indices (s, t) and dimension q , the corresponding persistent Betti numbers $\beta_q^{K^s, K^t}$ measure how many q -dimensional features are present at level K^s and still survive up to level K^t ; their formal definition is recalled in Section 3.6.2.

Persistent homology itself is purely topological: it only retains information encoded in the kernels of boundary operators, i.e. the null spaces, and disregards all non-zero spectral data. To enrich this picture, one may instead work with a refined operator that reproduces the same topological invariants while simultaneously capturing additional geometric structure in its non-zero spectrum [Wei and Wei, 2024, p. 10]. This operator is the *persistent Laplacian*, which will be introduced rigorously in Definition 4.2.4.

The strengths and limitations of the three underlying frameworks, classical homology, Hodge Laplacians, and persistent homology, are analysed systematically in Chapter 4, Section 4.1.

3.6.2 Filtrations in Practice

In applied settings, filtrations are often defined via proximity or threshold parameters.

Definition 3.6.1 (Simplicial filtration).

A *simplicial filtration* $\mathcal{K} = \{K_t\}_{t \in T}$ is a family of simplicial complexes on the same finite ordered vertex set V , indexed by $T \subseteq \mathbb{R}$, such that $K_s \hookrightarrow K_t$ for all $s \leq t$ in T .

Volume-based filtration. Let (G_{sym}, w) be the symmetrised transaction graph with non-negative edge weights $w(e)$. Fix a finite decreasing sequence of volume thresholds $v_0 > v_1 > \dots > v_M \geq 0$ and define

$$G_{v_k} := (V, \{e \in E \mid w(e) \geq v_k\}),$$

with clique complexes $K_{v_k} := K(G_{v_k})$. As v_k decreases, more edges are admitted and the complexes grow: if $v_k > v_\ell$, then $K_{v_k} \subseteq K_{v_\ell}$. To fit this family into the standard simplicial filtration framework of Definition 3.6.1, we introduce a monotone reparametrisation $t \mapsto v(t)$, for instance $t_k := -v_k$ or $t_k := k$, and write the filtration as $\{K_t\}_{t \in T}$ with $t_0 < t_1 < \dots < t_M$ and

$$t_1 \leq t_2 \Rightarrow K_{t_1} \subseteq K_{t_2}.$$

In particular, decreasing the volume threshold corresponds to increasing the filtration parameter t .

In transaction networks, filtrations arise naturally in two ways that are directly relevant for AML detection:

- *Temporal filtration*: the complex K_t is built from all transactions up to time t ; increasing t adds new edges and higher simplices as accounts interact over time.
- *Volume-based filtration*: the complex K_θ is the clique complex of the subgraph of G_{sym} retaining only edges with symmetrised weight $w_{\text{sym}}(e) \geq \theta$; decreasing θ gradually admits lower-volume transactions, making smurfing patterns visible as new 1-cycles that appear once θ drops to or below the individual transaction amounts. The non-strict inequality $\geq$ is adopted throughout this thesis (see Definition 4.7.3 and Remark 4.7.4).

Concrete constructions on AMLworld data are given in Chapter 5.

3.6.3 Persistent Homology Groups and Betti Numbers

Definition 3.6.2 (Persistent homology group of a filtration).

For any $s \leq t$ in T , functoriality of homology yields a linear map $f_q^{s,t} : H_q(K_s) \rightarrow H_q(K_t)$. The q -th *persistent homology group* of the pair (K_s, K_t) is $\text{im}(f_q^{s,t})$, and the q -th *persistent Betti number* $\beta_q^{s,t}$ is its rank.

Remark 3.6.3 (Notation: filtration pair vs. simplicial pair). Whenever the filtration pair (K_s, K_t) of Definition 3.6.2 is concretely given by a simplicial inclusion $K \subseteq L$, we write

$$\beta_q^{K,L} := \beta_q^{s,t} \quad (K = K_s, L = K_t).$$

Both notations are used interchangeably throughout Chapter 4; the pair-based notation $\beta_q^{K,L}$ is formalised in Definition 3.6.4.

Definition 3.6.4 (Persistent homology group of a simplicial pair).

Let $K \subseteq L$ be simplicial complexes and $q \geq 0$. The q -th persistent homology group of the pair is

$$H_q^{K,L} := \ker \partial_q^K / (\text{im } \partial_{q+1}^L \cap C_q(K)),$$

and its dimension is the q -th persistent Betti number of (K, L) .

In the volume filtration of Definition 4.7.3, the pair $(K, L) = (K_{\theta_s}, K_{\theta_{s+1}})$ arises from the convention $w_{\text{sym}}(e) \geq \theta$; see Remark 4.7.4. In particular, $\beta_q^{K,L} = \beta_q^{s,t}$ in the sense of Definition 3.6.2 whenever $K = K_s$ and $L = K_t$ (cf. Remark 3.6.3).

Lemma 3.6.5 (Image of the restricted boundary operator). *Let $K \subseteq L$ and $C_{q+1}^{K,L}, \partial_{q+1}^{K,L}$ as in Definition 3.6.6. Then*

$$\text{im } \partial_{q+1}^{K,L} = \text{im } \partial_{q+1}^L \cap C_q(K). \quad (3.3)$$

Proof. By definition of $C_{q+1}^{K,L}$, a chain $c \in C_{q+1}(L)$ belongs to $C_{q+1}^{K,L}$ if and only if $\partial_{q+1}^L(c) \in C_q(K)$. Therefore,

$$\begin{aligned} \text{im } \partial_{q+1}^{K,L} &= \{ \partial_{q+1}^L(c) : c \in C_{q+1}^{K,L} \} \\ &= \{ \partial_{q+1}^L(c) : c \in C_{q+1}(L), \partial_{q+1}^L(c) \in C_q(K) \} \\ &= \text{im } \partial_{q+1}^L \cap C_q(K). \end{aligned} \quad \square$$

Lemma 3.6.5 shows that the three equivalent formulations used in this thesis, namely $\text{im } \partial_{q+1}^{K,L}$, $\text{im } \partial_{q+1}^L \cap C_q(K)$, and the explicit quotient in Definition 3.6.4, all refer to the same subspace of $C_q(K)$. The persistent homology group may therefore be written equivalently as

$$H_q^{K,L} = \ker \partial_q^K / \text{im } \partial_{q+1}^{K,L} = \ker \partial_q^K / (\text{im } \partial_{q+1}^L \cap C_q(K)), \quad (3.4)$$

and both notations are used interchangeably in Chapter 4.

From a topological perspective, $\beta_q^{K,L}$ counts the q -dimensional homology classes that are “born” in K and remain “alive” in L , that is, cycles that have not yet been filled by higher-dimensional simplices added in L .

In the AML context, a persistent 1-cycle, corresponding to a high $\beta_1^{K,L}$, signals a fraud ring that remains topologically present across the filtration range from K to L [Wei and Wei, 2024, p. 9].

3.6.4 Definition of Persistent Laplacians

Persistent Laplacians extend the Hodge Laplacians of Section 3.5 from single complexes to simplicial pairs (K, L) and provide a spectral counterpart to persistent homology. Their definition is given below. The main structural results, including the Betti–nullity correspondence, are proved in Chapter 4.

Definition 3.6.6 (Persistent Laplacian on a simplicial pair).

Let $K \subseteq L$ be simplicial complexes and $q \geq 0$. Define

$$C_{q+1}^{K,L} := \{c \in C_{q+1}(L) : \partial_{q+1}^L(c) \in C_q(K)\},$$

and let $\partial_{q+1}^{K,L}$ be the restriction of ∂_{q+1}^L to $C_{q+1}^{K,L}$. The q -th persistent Laplacian is

$$\Delta_q^{K,L} : C_q(K) \rightarrow C_q(K), \quad \Delta_q^{K,L} := \underbrace{\partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^*}_{\Delta_{q,+}^{K,L}} + \underbrace{(\partial_q^K)^* \partial_q^K}_{\Delta_{q,-}^K}.$$

When $K = L$, this reduces to the standard Hodge q -Laplacian Δ_q^K .

Remark 3.6.7 (Preview notation). In this section we consider a pair of simplicial complexes $K \subseteq L$ to preview the persistent Laplacian $\Delta_q^{K,L}$. The full operator-theoretic treatment, including the Betti–nullity correspondence and Schur-complement representation for $\Delta_q^{K,L}$, is given in Chapter 4; see in particular Definition 4.2.4 and Theorem 4.2.10.

Theorem 3.6.8 (Persistent Hodge theorem). [*Wei and Wei, 2024*, pp. 10–11] *Let $K \subseteq L$ be simplicial complexes and $q \geq 0$. Then*

$$\ker \Delta_q^{K,L} \cong H_q^{K,L},$$

and $C_q(K)$ admits the orthogonal decomposition

$$C_q(K) = \text{im } \Delta_{q,+}^{K,L} \oplus \ker \Delta_q^{K,L} \oplus \text{im } \Delta_{q,-}^K.$$

The proof follows Steps 1–3 in Theorem 4.2.10; see also Remark 3.6.9 for a comparison of the two notations. Self-adjointness and positive semidefiniteness of $\Delta_q^{K,L}$ are established in Proposition 4.2.7.

3.6.5 Matrix Representation and Spectral Information

For numerical work, persistent Laplacians are assembled from boundary matrices. The key computational tool is a Schur-complement reduction of the full boundary matrix ∂_{q+1}^L , which yields a compact formula entirely in terms of the simplices of K . In the unweighted case ($w \equiv 1$), this reads

$$\Delta_q^{K,L} = \Delta_q^K + B_{12} B_{22}^\dagger B_{12}^\top,$$

where B_{12} and B_{22} are off-diagonal blocks of ∂_{q+1}^L indexed by S_q^K and $S_q^{L \setminus K}$, respectively; this follows from the Schur-complement identity for the up-persistent Laplacian [Mémoli et al., 2022, Theorem 4.6, Lemma 4.7].

The full derivation for the weighted case and two explicit algorithms are presented in Section 4.3.2.

Remark 3.6.9. The zero eigenvalues of $\Delta_q^{K,L}$ recover the persistent Betti numbers [Mémoli et al., 2022, Theorem 2.7], while the non-zero eigenvalues encode additional geometric information about the filtration, such as the scales at which features are weakly or strongly represented, that is invisible to homology alone [Wei and Wei, 2024, pp. 4, 10–11]. This makes the persistent Laplacian a strictly finer invariant than persistent homology: two filtration pairs with the same persistent Betti number but different geometry will generically have different non-zero spectra [Wei and Wei, 2024, pp. 10–11].

3.6.6 Example: Transaction Ring in a Filtration

Example. Consider the transaction network from Example 3.1.2 with an amount-based filtration obtained by lowering a transaction-volume threshold. At high thresholds, only edges with $w_{\text{sym}}(e) \geq \theta$ are retained; in the ring $A \rightarrow B \rightarrow C \rightarrow A$ these are precisely the high-volume arcs, so the corresponding 1-dimensional homology class is born early and persists over a wide range of θ . The low-volume edges $A \rightarrow D$ and $D \rightarrow C$ satisfy $w_{\text{sym}}(e) \geq \theta$ only once θ drops to or below their individual weights, and generate at most short-lived peripheral features.

From the viewpoint of persistent Laplacians (Remark 3.6.9), the kernel of $\Delta_1^{K,L}$ reflects the persistent ring $A-B-C$ [Mémoli et al., 2022, p. 7, Theorem 2.7], while the non-zero eigenvalues encode its geometric strength at each filtration scale.

Remark 3.6.10 (Loops and persistent Laplacians). As established in Remark 3.1.2, loops are excluded from all spectral and topological constructions. In the filtration setting this is automatically ensured: the clique complexes $K_t = K(G_{\text{sym}}^t)$ are built from the symmetrised graph at each filtration step, and 1-simplices are two-element subsets $\{u, v\}$ with $u \neq v$ by definition. Self-transfers, if present in the raw data, therefore never contribute to the persistent Laplacian or to the persistent Betti numbers $\beta_1^{K,L}$.

Chapter 4 develops the full operator-theoretic and algorithmic machinery: the Betti-nullity correspondence (Theorem 4.2.10) closes the loop between the spectral and homological perspectives, the Schur-complement formula (Theorem 4.3.3) makes eigenvalue computation tractable, and the persistent Cheeger inequality (Theorem 4.4.9) connects the spectrum directly to the layering-bottleneck interpretation of Section 2.3.3.

Chapter 4

Persistent Laplacians for Transaction Network Analysis

4.1 From Persistent Homology to the Persistent Laplacian

This chapter develops the theory and algorithmics of the persistent Laplacian for AML transaction networks and formulates the structural guarantees (monotonicity, stability, Cheeger-type bounds) that are numerically put to the test in Chapter 5 on the pattern-centred working graph G_{work} .

4.1.1 Three Operators, Three Limitations

The analysis of cyclic transaction structures requires an operator that simultaneously meets three structural criteria: it must track topological features *across filtration scales* (Definition 3.6.1), it must distinguish structures that are identical up to homotopy equivalence but differ in their geometric or quantitative realization, in line with the limitations of persistent homology discussed in Wei and Wei [2024, p. 3], and it must act on a concrete, computable function space. No classical operator simultaneously fulfils all three requirements in a unified framework. This observation motivates the construction of the persistent Laplacian and clarifies the respective limitations of persistent homology and classical Hodge-theoretic operators.

The five graph-theoretic AML patterns identified in Table 2.1 of Chapter 2, namely cycles and rings, fan-in/fan-out structures, layering cascades, strongly connected components, and temporal bursts, each leave a distinct signature in the topology of the transaction complex. Table 4.1 below indicates which of the three operator classes can capture each type of signature; the precise correspondence is given in Table 4.2 of Section 4.8.

Persistent homology. Given a filtration $\mathcal{K} = \{K_t\}_{t \in T}$ in the sense of Definition 3.6.1, persistent homology records, for each homology class, a birth–death interval $[b, d)$; see Section 3.6.3 for the precise construction. The resulting barcode is a homotopy invariant: two simplicial complexes with the same homotopy type but different finer metric or geometric properties are indistinguishable by their barcodes [Wei and Wei, 2024, pp. 3–4].

Moreover, the barcode records only the *dimension* of the persistent homology groups, that is, the number of surviving cycles, but carries no information about their geometric weight or internal structure [Wei and Wei, 2024, p. 3].

Remark 4.1.1. In the AML context this limitation is directly operational. A compact, high-volume fraud ring (Section 2.3.1) and a loose, sparse cycle of benign transactions may carry identical one-dimensional persistent bars if their topological type coincides. Persistent homology alone cannot separate them. From a spectral viewpoint, persistent Laplacians attach, at each filtration step, a self-adjoint, positive semi-definite Laplace operator whose kernel is naturally isomorphic to the corresponding persistent homology group and whose nullity equals the persistent Betti number [Mémoli et al., 2022, Thm. 2.7, Sec. 2.3][Liu et al., 2024, Cor. 3.9, Thm. 3.10]. The ambient chain space admits an orthogonal decomposition into this harmonic subspace and two Laplacian images, so that the non-zero spectrum of $\Delta_q^{K,L}$ resides entirely on complementary components and thereby carries information about the filtration that is not contained in the barcodes alone. In practical terms, this means that the first Betti number β_1^K of a single complex (Section 3.4.4) detects the *existence* of a cyclic structure, whereas the persistent spectral gap $\lambda_{\text{gap}}^{(1)}(K, L)$ (Definition 4.8.1) is a quantitative measure of the geometric robustness of such a cyclic structure across a filtration step. This topological–spectral distinction is critical for separating structurally strong fraud rings from transient noise.

Classical graph and Hodge Laplacians. The combinatorial Laplacian Δ_q^K , introduced in Definition 3.5.5, is a self-adjoint, positive semi-definite operator on $C_q(K)$ whose nullity equals the q -th Betti number β_q^K , as shown in Theorem 3.5.11. For $q = 0$, it reduces to the classical graph Laplacian [Mémoli et al., 2022, p. 5]; its Fiedler value $\lambda_1(L_{\text{sym}})$, the smallest positive eigenvalue of the symmetrised Laplacian L_{sym} in the sense of Definition 3.3.10, provides a quantitative measure of connectivity via Cheeger-type inequalities [Lim, 2020, p. 17][Mémoli et al., 2022, Thm. 4.20], and $\beta_1^K = \dim \ker(\Delta_1^K)$ counts independent cyclic flow patterns (Section 3.4.4). Both quantities already appear as spectral indicators in Table 2.1: $\lambda_1(L_{\text{sym}})$ signals layering cascades and strongly connected components (Sections 2.3.3–2.3.4), while $\beta_1^K \geq 1$ signals cycles and rings (Section 2.3.1).

However, Δ_q^K operates at a fixed scale: it is defined on a single simplicial complex K and contains no filtration parameter, so it captures only a snapshot of the network at a given transaction threshold. Consequently, it carries no information about how structure evolves as the threshold varies and is therefore insensitive to the multi-scale persistence of fraud-ring patterns that characterise layering cascades.

The persistent Laplacian as the natural synthesis. The persistent q -Laplacian $\Delta_q^{K,L}$, defined formally in Section 4.2, resolves both deficiencies simultaneously. It acts on the chain group $C_q(K)$ of the smaller complex while encoding the ambient structure of L through the restricted boundary operator $\partial_{q+1}^{K,L}$; the inner product conventions are those fixed in Remark 3.5.2.

As a self-adjoint, positive semi-definite operator on a finite-dimensional inner product space [Liu et al., 2024, p. 11, Rem. 3.3], it admits a complete real spectral decomposition and the persistence spectra vary continuously under perturbations of the filtration, as quantified by the stability theorem [Mémoli et al., 2022, Thm. 5.9].

By the persistent Hodge theorem (Theorem 3.6.8), $C_q(K)$ admits the orthogonal decomposition (4.11) established in Section 4.2.3 [Liu et al., 2024, Thm. 3.10].

Table 4.1 summarises the structural capabilities of the three operators discussed above.

Table 4.1: Comparison of spectral and topological operators for transaction network analysis. The AML pattern column refers to the typologies formalised in Table 2.1.

	Persistent Homology	Graph / Hodge Laplacian	Persistent Laplacian	AML pattern (Table 2.1)
Multi-scale (filtration-aware)	✓	×	✓	Layering cascades, temporal bursts
Higher-order struc- ture ($q \geq 1$)	✓	✓	✓	Cycles, SCCs
Geometric / metric information	×	✓	✓	Distinguishes fraud rings from benign cycles
Self-adjoint opera- tor on $C_q(K)$	×	✓	✓	—
PSD, real spec- trum	×	✓	✓	—
Recovers Betti numbers β_q^K	✓	✓	✓	Cycles ($\beta_1^K \geq 1$), SCCs ($\beta_0^K > 1$)
Recovers persistent Betti numbers $\beta_q^{K,L}$	✓	×	✓	Persistent fraud rings, temporal bursts
Stable under filtra- tion perturbations	✓	×	✓	Robust to volume noise in smurfing de- tection

Having compared the three operators, we now define the persistent Laplacian formally and establish its fundamental properties.

4.1.2 Operator Structure and Key Properties

The persistent q -Laplacian decomposes as

$$\Delta_q^{K,L} = \underbrace{\partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^*}_{\Delta_{q,+}^{K,L}, \text{ filtration-sensitive}} + \underbrace{(\partial_q^K)^* \partial_q^K}_{\Delta_{q,-}^K, \text{ depends only on } K}, \quad (4.1)$$

where $\partial_{q+1}^{K,L} : C_{q+1}^{K,L} \rightarrow C_q(K)$ is the restriction of the full boundary operator ∂_{q+1}^L to

$$C_{q+1}^{K,L} := \{ c \in C_{q+1}(L) \mid \partial_{q+1}^L c \in C_q(K) \},$$

as defined in Definition 3.6.6.

Figure 4.1 summarizes the operator structure of the persistent q -Laplacian. The diagonal blue arrows $\partial_{q+1}^{K,L}$ and $(\partial_{q+1}^{K,L})^*$ encode the filtration-sensitive up-persistent part $\Delta_{q,+}^{K,L}$, whereas the horizontal blue arrows in the top row represent the down-Laplacian $\Delta_{q,-}^K$ acting purely on $C_q(K)$.

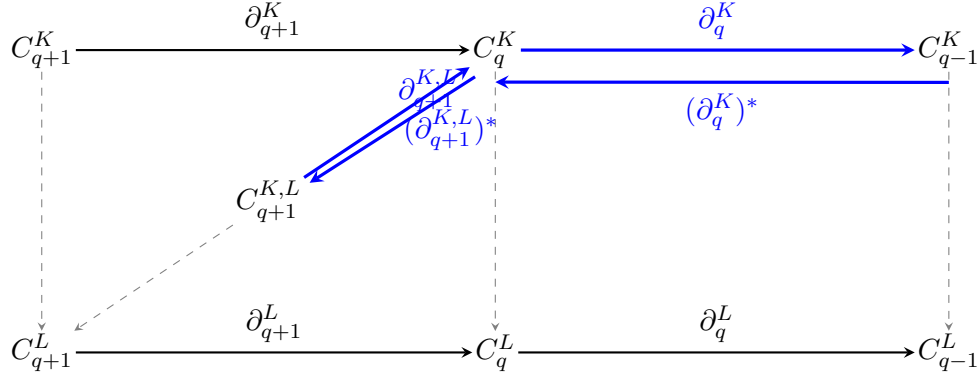


Figure 4.1: Illustration of the up-persistent Laplacian $\Delta_{q,+}^{K,L}$ and down persistent Laplacian $\Delta_{q,-}^K$ along a filtration $K \hookrightarrow L$. Inspired by [Mémoli et al. \[2022, p. 6\]](#).

The two summands satisfy a mixed-orthogonality relation. Indeed,

$$\Delta_{q,+}^{K,L} \circ \Delta_{q,-}^K = \partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^* (\partial_q^K)^* \partial_q^K.$$

Since $\partial_q^K \circ \partial_{q+1}^{K,L} = 0$ by the boundary-of-a-boundary identity (Definition 3.4.7), it follows that

$$(\partial_{q+1}^{K,L})^* (\partial_q^K)^* = (\partial_q^K \circ \partial_{q+1}^{K,L})^* = 0,$$

which yields

$$\Delta_{q,+}^{K,L} \circ \Delta_{q,-}^K = 0.$$

In particular, the images of the up- and down-Laplacians are orthogonal,

$$\text{Im}(\Delta_{q,+}^{K,L}) \perp \text{Im}(\Delta_{q,-}^K),$$

in accordance with the persistent Hodge decomposition (4.11). As a consequence, since the condition $BA = 0$ implies that the non-zero spectrum of the joint operator $AA^* + B^*B$ decomposes into the disjoint union of the individual spectra of AA^* and B^*B , the non-zero spectrum of $\Delta_q^{K,L}$ is precisely the disjoint union, with multiplicity, of the non-zero spectra of $\Delta_{q,+}^{K,L}$ and $\Delta_{q,-}^K$ [Botnan and Dong, 2025, p. 21].

The kernel of $\Delta_q^{K,L}$ is exactly the persistent harmonic space; by Theorem 3.6.8 it is naturally isomorphic to the q -th persistent homology group [Liu et al., 2024, p. 12, Cor. 3.9, Thm. 3.10] [Mémoli et al., 2022, p. 7, Thm. 2.7]:

$$\dim \ker \Delta_q^{K,L} = \beta_q^{K,L}, \quad (4.2)$$

where $\beta_q^{K,L}$ is the q -th persistent Betti number of the pair (K, L) in the sense of Definition 3.6.4.

For AML detection, the three key properties are the following.

1. **Kernel = persistent homology.** By (4.2) and Theorem 4.2.10, every fraud ring detected by persistent homology is reflected in the kernel of $\Delta_q^{K,L}$; no topological information is lost. In terms of Table 2.1, the signal $\beta_1 \geq 1$ for cycles and rings (Section 2.3.1) is fully preserved as $\beta_1^{K,L} = \text{nullity}(\Delta_1^{K,L})$.
2. **Non-zero spectrum = geometric refinement.** By the persistent <hodge Theorem, the kernel of $\Delta_q^{K,L}$ encodes precisely the persistent Betti numbers, and the non-zero eigenvalues provide strictly richer information about the filtration than any barcode invariant alone [Wei and Wei, 2024, pp. 10–11]. In the AML model, the non-harmonic eigenvalues are used to quantify how strongly cyclic transaction patterns are geometrically realised at a given filtration scale. The persistent spectral gap $\lambda_{\text{gap}}^{(1)}(K, L)$ serves as a summary statistic of this intensity.
3. **Algebraic stability.** For two filtration functions f, g with $\|f - g\|_\infty \leq \varepsilon$, the interleaving distance between the resulting persistent Laplacian spectra satisfies

$$d_I(\mathcal{L}_f, \mathcal{L}_g) \leq \|f - g\|_\infty \leq \varepsilon \quad (4.3)$$

[Liu et al., 2024, p. 20, Thm. 5.3], guaranteeing that AML-detection signatures are robust to small perturbations in transaction volumes or timestamps. The full statement and proof are given in Section 4.4; the practical implications of this stability bound for the chosen window length and volume thresholds, and the limitations of the current parameter calibration on the

AMLWORLD dataset, are discussed in Section 5.6.4. In particular, the theoretical bound (4.3) implies that perturbations below the smallest observed edge-weight increment which do not change which edges satisfy $w_{\text{sym}}(e) \geq \theta$ for any threshold θ in the sequence $\theta_0 > \dots > \theta_T$ leave the Laplacian-tree feature vector unchanged up to numerical precision. This is critical for the smurfing and fan-in/fan-out scenarios of Section 2.3.2, where transactions are deliberately kept close to reporting thresholds and small day-to-day variations in amounts or timestamps should not qualitatively change the detected spectral-topological signature.

With respect to the research questions formulated in Chapter 1, Chapter 4 thus addresses the core problem of money-laundering detection under extreme class imbalance: it constructs persistent spectral operators whose kernels capture the rare, structurally relevant fraud rings and layering cascades in a topological sense, while their non-zero spectra provide robust, interpretable measures of geometric intensity. These quantities are aggregated into window- and account-level feature vectors on G_{work} in Chapter 5 and empirically evaluated against the fraud labels of the AMLWORLD dataset.

Remark 4.1.2 (Regulatory interpretability). Under Article 8(2) of Directive (EU) 2015/849 [European Parliament and Council, 2015, Art. 8(2)], obliged entities must document and justify their risk-assessment methodology, as mentioned in Section 2.1.3. The features derived from $\Delta_q^{K,L}$ satisfy this requirement: $\beta_1^{K,L}$ counts independent cycles (Definition 3.6.4), the persistent spectral gap $\lambda_{\text{gap}}^{(1)}(K, L)$ measures their geometric strength (Definition 4.8.1), and the spectral entropy H_{spec} (Section 4.6) quantifies distributional complexity.

Each quantity is directly relatable to an AML typology via Table 4.2, and is therefore documentable under regulatory audit, in contrast to the representations of deep neural networks [Deprez et al., 2025b, pp. 28–29].

4.2 Definition and Fundamental Properties

Remark 4.2.1. Throughout Chapter 4 we write $\Delta_q^{K,L}$ for the persistent Laplacian of the pair ($K \subseteq L$) and Δ_q^K for the standard Hodge Laplacian of a single complex K . All chain groups in this chapter carry the *canonical (unweighted) ℓ^2 inner product* with respect to their simplex bases. Weighted inner products from Section 3.5.4 are not used in the proofs of this chapter.

4.2.1 Definition via the Restricted Boundary Operator

Let $K \subseteq L$ be a simplicial pair in the sense of Definition 3.6.6, and fix $q \geq 0$. The key construction is a subspace of $(q+1)$ -chains in L whose boundary already lies in K .

Definition 4.2.2 (Up-persistent chain group and restricted boundary operator). The *up-persistent $(q+1)$ -chain group* of the pair (K, L) is

$$C_{q+1}^{K,L} := \{ c \in C_{q+1}(L) : \partial_{q+1}^L(c) \in C_q(K) \},$$

and $\partial_{q+1}^{K,L}: C_{q+1}^{K,L} \rightarrow C_q(K)$ denotes the restriction of ∂_{q+1}^L to $C_{q+1}^{K,L}$.

Remark 4.2.3 (Adjoint vs. Moore–Penrose pseudoinverse). Throughout this chapter, A^* denotes the *Hilbert-space adjoint* of a linear map A with respect to the canonical ℓ^2 -inner products on domain and codomain; in particular, $(\partial_{q+1}^{K,L})^*: C_q(K) \rightarrow C_{q+1}^{K,L}$ is represented in the ordered simplex bases by the matrix transpose $B^\top$ (since both spaces carry the unweighted ℓ^2 -inner product, cf. Remark 4.2.1). The symbol $A^\dagger$ denotes *exclusively* the Moore–Penrose pseudoinverse of a matrix, which appears in the algorithmic Schur-complement formula of Theorem 4.3.3 as $B_{22}^\dagger$. These two notations are never conflated: $A^* = A^\top$ whenever A^* appears in proofs, and $A^\dagger$ whenever a numerical least-squares inverse is required. (In Chapter 3 the adjoint of a boundary operator is occasionally written $\partial^\dagger$ following Lim [2020, p. 30], within Chapter 3 this refers to the same object as ∂^* here.)

Since $K \subseteq L$, the chain inclusions $C_{q+1}(K) \subseteq C_{q+1}^{K,L} \subseteq C_{q+1}(L)$ hold. The group $C_{q+1}^{K,L}$ consists precisely of those $(q+1)$ -chains of L whose boundary is entirely supported on K [Mémoli et al., 2022, p. 6].

Equipping $C_{q+1}^{K,L}$ and $C_q(K)$ with the canonical ℓ^2 inner product inherited from $C_{q+1}(L)$ and $C_q(L)$, respectively (Remark 4.2.1), the restricted boundary operator $\partial_{q+1}^{K,L}$ of Definition 4.2.2 admits a unique Hilbert-space adjoint

$$(\partial_{q+1}^{K,L})^*: C_q(K) \longrightarrow C_{q+1}^{K,L},$$

following the construction of Botnan and Dong [2025, p. 10, Def. 3.3].

Definition 4.2.4 (Persistent q -Laplacian). The q -th persistent Laplacian of the simplicial pair $(K \subseteq L)$ is the linear map

$$\Delta_q^{K,L} := \underbrace{\partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^*}_{\Delta_{q,+}^{K,L} \text{ (up part)}} + \underbrace{(\partial_q^K)^* \partial_q^K}_{\Delta_{q,-}^K \text{ (down part)}} : C_q(K) \longrightarrow C_q(K). \quad (4.4)$$

When $K = L$, both summands reduce to the corresponding summands of the standard Hodge q -Laplacian Δ_q^K of Definition 3.5.5.

Note that the down part $\Delta_{q,-}^K$ carries only the superscript K : it depends solely on the inner complex K and is independent of the ambient complex L . The filtration-sensitive dependence on L is carried exclusively by $\Delta_{q,+}^{K,L}$ via the restricted operator $\partial_{q+1}^{K,L}$ of Definition 4.2.2.

Lemma 4.2.5. *Let $K \subseteq L$ be a simplicial pair and $q \geq 0$. If $C_{q+1}^{K,L} = 0$, then $\partial_{q+1}^{K,L} = 0$ and consequently $\Delta_{q,+}^{K,L} = 0$, so that $\Delta_q^{K,L} = \Delta_{q,-}^K$.*

Proof. By Definition 4.2.2, $\partial_{q+1}^{K,L}$ is the restriction of ∂_{q+1}^L to $C_{q+1}^{K,L}$. If $C_{q+1}^{K,L} = \{0\}$, the operator acts on a trivial domain and is therefore the zero map. Hence $\Delta_{q,+}^{K,L} = \partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^* = 0$, [Mémoli et al., 2022, p. 6]. $\square$

Remark 4.2.6. The two summands capture complementary information. The down part $\Delta_{q,-}^K = (\partial_q^K)^* \partial_q^K$ depends only on K and coincides with the down-Laplacian of K alone. The up part $\Delta_{q,+}^{K,L} = \partial_{q+1}^{K,L} (\partial_{q+1}^{K,L})^*$ is the *filtration-sensitive* summand: it detects $(q+1)$ -chains in L whose boundary reaches into K , and it vanishes whenever $C_{q+1}^{K,L} = 0$ (Lemma 4.2.5). In the AML context, $\Delta_{q,+}^{K,L}$ reflects how simplices entering at the next filtration step may close cycles already present in K , which is precisely the mechanism by which layering transactions (Section 2.3.3) build up cyclic structures incrementally.

4.2.2 Self-Adjointness and Positive Semidefiniteness

Proposition 4.2.7. *For any simplicial pair $K \subseteq L$ and any $q \geq 0$, where all chain groups carry the ℓ^2 inner product with respect to their simplex bases (Remark 4.2.1):*

- (1) $\Delta_q^{K,L}$ is self-adjoint with respect to the ℓ^2 inner product on $C_q(K)$.
- (2) $\Delta_q^{K,L}$ is positive semidefinite: $\langle \Delta_q^{K,L} c, c \rangle \geq 0$ for all $c \in C_q(K)$.
- (3) The eigenvalues of $\Delta_q^{K,L}$ are non-negative real numbers; denoted in non-decreasing order,

$$0 \leq \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n_q^K - 1},$$

each counted with multiplicity, where $n_q^K := \dim C_q(K)$.

Proof. (1) Each summand is of the form AA^* (for $\Delta_{q,+}^{K,L}$) respectively B^*B (for $\Delta_{q,-}^K$). We verify self-adjointness explicitly in the present setting. The chain group $C_q(K)$ is a finite-dimensional real vector space with ordered simplex basis $\{\sigma_1, \dots, \sigma_{n_q^K}\}$, equipped with the ℓ^2 inner product $\langle c, d \rangle = \sum_{i=1}^{n_q^K} c_i d_i$ (Remark 4.2.1). The restricted boundary operator $\partial_{q+1}^{K,L}$ is represented in this basis by a real matrix $B \in \mathbb{R}^{n_q^K \times m}$, where $m = \dim C_{q+1}^{K,L}$, and its adjoint $(\partial_{q+1}^{K,L})^*$ is represented by the transpose $B^\top$ (since the inner products on both spaces are the canonical ℓ^2 products). Therefore $\Delta_{q,+}^{K,L}$ is represented by $BB^\top$, which satisfies $(BB^\top)^\top = (B^\top)^\top B^\top = BB^\top$, and is hence symmetric. Analogously, $\Delta_{q,-}^K$ is represented by $C^\top C$ for the matrix C of ∂_q^K , which is likewise symmetric. The sum of two symmetric matrices is symmetric, so $\Delta_q^{K,L}$ is self-adjoint.

- (2) For any $c \in C_q(K)$, the decomposition (4.4) gives

$$\langle \Delta_q^{K,L} c, c \rangle = \langle \Delta_{q,+}^{K,L} c, c \rangle + \langle \Delta_{q,-}^K c, c \rangle = \|(\partial_{q+1}^{K,L})^* c\|^2 + \|\partial_q^K c\|^2 \geq 0, \quad (4.5)$$

where the second equality uses $\langle AA^* c, c \rangle = \|A^* c\|^2$ and $\langle B^* B c, c \rangle = \|B c\|^2$ for any linear maps A, B . Hence $\Delta_q^{K,L}$ is positive semi-definite. Moreover, $\langle \Delta_q^{K,L} c, c \rangle = 0$ if and only if both $(\partial_{q+1}^{K,L})^* c = 0$ and $\partial_q^K c = 0$ simultaneously, since each summand is a squared norm and therefore non-negative.

- (3) The claim on the spectrum follows from (1) and (2) by the spectral theorem for finite-dimensional real symmetric matrices (Theorem 3.2.18): every such matrix is diagonalizable with real eigenvalues, and positive semidefiniteness forces all eigenvalues to be non-negative. $\square$

This result is stated for general weighted inner products already in [Mémoli et al. \[2022, p. 6\]](#), and generalized categorically to arbitrary morphisms of differential graded inner product spaces in [Liu et al. \[2024, p. 11, Rem. 3.3\]](#).

Remark 4.2.8 (Scope of Proposition 4.2.7). In Sections 4.7–4.8.1, the edge weights w_{ij} of G_{sym} determine the order in which simplices enter the filtration $\mathcal{K} = \{K(\tau)\}_{\tau \in T}$ (Definition 3.6.1) but do *not* enter the chain-group inner product. All chain groups $C_q(K(\tau))$ carry the canonical unweighted ℓ^2 inner product throughout, in accordance with Remark 4.2.1. Proposition 4.2.7 therefore applies without modification to every persistent Laplacian $\Delta_q^{K(\tau), K(\tau')}$ appearing in those sections. Proposition 4.2.7 is stated and proved here for the unweighted case. The weighted generalisation, in which each chain group $C_q(K)$ carries a weight-induced inner product $\langle c, d \rangle_w = \sum_{\sigma} w(\sigma)^{-1} c_{\sigma} d_{\sigma}$ (Remark 3.5.2), is covered by [Mémoli et al. \[2022, Rem. 3.2\]](#), which shows that the persistent Laplacian takes the form $W^{-1}PW$ for a symmetric positive semi-definite matrix P and a positive diagonal weight matrix W . Self-adjointness and positive semi-definiteness with respect to the w -weighted inner product follow immediately.

4.2.3 The Betti–nullity correspondence

The following theorem is the algebraic cornerstone of the framework: it shows that the kernel of $\Delta_q^{K,L}$ is canonically isomorphic to the q -th persistent homology group of the pair.

Remark 4.2.9 (Origin and interrelation of the Betti–nullity correspondence). The identity established below instantiates, for the setting of simplicial pairs, a persistent version of the classical Eckmann–Hodge theorem $\beta_q(K) = \text{nullity}(\Delta_q^K)$ for a single simplicial complex K [[Wei and Wei, 2024](#), pp. 3–4, pp. 16–17]. [Mémoli et al. \[2022\]](#) gave the first fully detailed algebraic proof of the persistent version stated here ([[Mémoli et al., 2022](#), Thm. 2.7, Appendix A]), via the elementary fact that for matrices A, B with $AB = 0$, $\ker A \cap \text{im } B \cong \ker A \cap \ker B^T$, the same mechanism instantiated in Steps 1–2 of the proof below. [Liu et al. \[2024\]](#) generalized this correspondence categorically to arbitrary morphisms of differential graded inner product spaces [[Liu et al., 2024](#), Thm. 3.10], proving the abstract persistent Hodge isomorphism $\ker \Delta_q^{V,W} \cong \ker d_q^V / (\ker d_q^V \cap \text{im } d_{q+1}^W)$ [[Wei and Wei, 2024](#), Sec. 4.1], the simplicial statement of [Mémoli et al. \[2022\]](#) recovers as the special case where the morphism is the inclusion $C_q(K) \hookrightarrow C_q(L)$.

The proof given below follows the elementary, matrix-level approach of [Mémoli et al. \[2022\]](#) directly, since it yields the explicit basis and orthogonal-decomposition structure needed for the Schur-complement computations of Section 4.3.

Theorem 4.2.10 (Betti–nullity correspondence [Mémoli et al., 2022, p. 7, Thm. 2.7], [Wei and Wei, 2024, pp. 10–12], [Liu et al., 2024, p. 12, Cor. 3.9]). *Let $K \subseteq L$ be a simplicial pair and $q \geq 0$. Then*

$$\ker \Delta_q^{K,L} \cong H_q^{K,L} := \frac{\ker \partial_q^K}{\operatorname{im} \partial_{q+1}^{K,L}}, \quad (4.6)$$

where the denominator equals $\operatorname{im} \partial_{q+1}^L \cap C_q(K)$ by definition of $C_{q+1}^{K,L}$. In particular,

$$\operatorname{nullity}(\Delta_q^{K,L}) = \beta_q^{K,L} = \dim \ker \partial_q^K - \operatorname{rank}(\partial_{q+1}^{K,L}), \quad (4.7)$$

where $\beta_q^{K,L}$ is the persistent Betti number of the pair $K \subseteq L$ in the sense of Definition 3.6.4, equivalently $\beta_q^{s,t}$ of Definition 3.6.2 with $K = K_s$, $L = K_t$ (Remark 3.6.3).

Proof. We proceed in three steps.

Step 1: Kernel characterisation. By Proposition 4.2.7 (2), $\Delta_q^{K,L}c = 0$ if and only if $\langle \Delta_q^{K,L}c, c \rangle = 0$. From (4.5) this holds if and only if both

$$\partial_q^K c = 0 \quad \text{and} \quad (\partial_{q+1}^{K,L})^* c = 0$$

simultaneously, since both summands in (4.5) are non-negative and their sum vanishes only when each vanishes individually. Here $(\partial_{q+1}^{K,L})^*: C_q^K \rightarrow C_{q+1}^{K,L}$ denotes the adjoint of $\partial_{q+1}^{K,L}: C_{q+1}^{K,L} \rightarrow C_q^K$ with respect to the inner products. Hence

$$\ker \Delta_q^{K,L} = \ker \partial_q^K \cap \ker (\partial_{q+1}^{K,L})^*. \quad (4.8)$$

Step 2: Harmonic representatives. We first show that $\ker (\partial_{q+1}^{K,L})^* = (\operatorname{im} \partial_{q+1}^{K,L})^\perp$, where the orthogonal complement is taken inside the ambient space C_q^K . For any $c \in C_q^K$, the condition $(\partial_{q+1}^{K,L})^*c = 0$ holds if and only if $\langle (\partial_{q+1}^{K,L})^*c, \sigma \rangle = 0$ for all $\sigma \in C_{q+1}^{K,L}$. By definition of the adjoint, this is equivalent to $\langle c, \partial_{q+1}^{K,L}\sigma \rangle = 0$ for all $\sigma \in C_{q+1}^{K,L}$, i.e. $c \in (\operatorname{im} \partial_{q+1}^{K,L})^\perp$ (complement in C_q^K). Thus

$$\ker (\partial_{q+1}^{K,L})^* = (\operatorname{im} \partial_{q+1}^{K,L})^\perp_{C_q^K}.$$

Next we record the inclusion needed for the orthogonal splitting below. Since $\partial_q^K \circ \partial_{q+1}^{K,L} = 0$ (the composite of two consecutive boundary maps in the pair complex), we have

$$\operatorname{im} \partial_{q+1}^{K,L} \subseteq \ker \partial_q^K. \quad (4.9)$$

Because $\ker \partial_q^K$ is a finite-dimensional real inner product space (as a subspace of the finite-dimensional space C_q^K) and $\operatorname{im} \partial_{q+1}^{K,L}$ is a subspace of it by (4.9), the orthogonal decomposition theorem for finite-dimensional inner product spaces gives

$$\ker \partial_q^K = \operatorname{im} \partial_{q+1}^{K,L} \oplus (\operatorname{im} \partial_{q+1}^{K,L})^\perp_{\ker \partial_q^K},$$

where the complement is now (equivalently, by (4.9)) taken either inside $\ker \partial_q^K$ or inside C_q^K — both descriptions agree because $\ker \partial_q^K \subseteq C_q^K$ already contains $\text{im } \partial_{q+1}^{K,L}$, so no vector of $C_q^K \setminus \ker \partial_q^K$ can contribute to either complement. Hence every $c \in \ker \partial_q^K$ decomposes uniquely as

$$c = c_{\text{harm}} + c_{\text{exact}}, \quad c_{\text{harm}} \in (\text{im } \partial_{q+1}^{K,L})^\perp \cap \ker \partial_q^K, \quad c_{\text{exact}} \in \text{im } \partial_{q+1}^{K,L}. \quad (4.10)$$

We now show that the assignment $\Phi: H_q^{K,L} \rightarrow \ker \Delta_q^{K,L}$, $[c] \mapsto c_{\text{harm}}$, is a well-defined linear isomorphism.

Well-definedness. Suppose $c, c' \in \ker \partial_q^K$ represent the same class, $[c] = [c']$, i.e. $c - c' \in \text{im } \partial_{q+1}^{K,L}$. Applying the (unique) decomposition (4.10) to $c - c'$ shows $(c - c')_{\text{harm}} = 0$, and by linearity of the projection onto the orthogonal summand, $(c - c')_{\text{harm}} = c_{\text{harm}} - c'_{\text{harm}}$. Hence $c_{\text{harm}} = c'_{\text{harm}}$, so Φ does not depend on the choice of representative.

$c_{\text{harm}} \in \ker \Delta_q^{K,L}$. Since $\ker \partial_q^K$ is a subspace containing both c and c_{exact} , it also contains $c_{\text{harm}} = c - c_{\text{exact}}$, so $\partial_q^K c_{\text{harm}} = 0$. Moreover $c_{\text{harm}} \in (\text{im } \partial_{q+1}^{K,L})^\perp = \ker(\partial_{q+1}^{K,L})^*$, so $(\partial_{q+1}^{K,L})^* c_{\text{harm}} = 0$. By (4.8), $c_{\text{harm}} \in \ker \Delta_q^{K,L}$, so Φ is well-defined as a map into $\ker \Delta_q^{K,L}$.

Injectivity. Suppose $\Phi([c]) = \Phi([c'])$, i.e. $c_{\text{harm}} = c'_{\text{harm}}$. By (4.10), $c - c' = c_{\text{exact}} - c'_{\text{exact}} \in \text{im } \partial_{q+1}^{K,L}$, hence $[c] = [c']$ in $H_q^{K,L}$. Thus Φ is injective.

Surjectivity. Let $h \in \ker \Delta_q^{K,L}$. By (4.8), $\partial_q^K h = 0$ and $(\partial_{q+1}^{K,L})^* h = 0$, so $h \in \ker \partial_q^K$ and $h \in (\text{im } \partial_{q+1}^{K,L})^\perp$. The decomposition (4.10) applied to h then gives $h_{\text{harm}} = h$ and $h_{\text{exact}} = 0$, i.e. $\Phi([h]) = h$. Thus Φ is surjective.

Since Φ is clearly linear (the orthogonal projection onto $(\text{im } \partial_{q+1}^{K,L})^\perp \cap \ker \partial_q^K$ is linear), Φ is a linear isomorphism $H_q^{K,L} \xrightarrow{\sim} \ker \Delta_q^{K,L}$, proving (4.6).

Step 3: Dimension of the quotient space. By (4.9), the quotient $\ker \partial_q^K / \text{im } \partial_{q+1}^{K,L} = H_q^{K,L}$ is well-defined as a real vector space. The standard dimension formula for quotients of finite-dimensional vector spaces then yields

$$\dim H_q^{K,L} = \dim \ker \partial_q^K - \dim \text{im } \partial_{q+1}^{K,L} = \dim \ker \partial_q^K - \text{rank}(\partial_{q+1}^{K,L}).$$

Combining this with $\dim \ker \Delta_q^{K,L} = \dim H_q^{K,L}$ (Step 2) gives

$$\text{nullity}(\Delta_q^{K,L}) = \dim \ker \partial_q^K - \text{rank}(\partial_{q+1}^{K,L}) = \beta_q^{K,L},$$

which proves (4.7). □

The Betti–nullity correspondence enables a Schur-complement computation of the full persistent spectrum, which is the subject of the next Subsection 4.3.

Remark 4.2.11. When $K = L$, equation (4.6) reduces to the classical combinatorial Hodge theorem $\ker \Delta_q^K \cong H_q(K)$ of Theorem 3.5.11. The persistent Laplacian thus strictly generalises the Hodge Laplacian of Definition 3.5.6.

Remark 4.2.12. For $q = 0$, the up part $\Delta_{0,+}^{K,L}$ coincides with the Kron reduction of the graph Laplacian of L to the vertex set of K [Mémoli et al., 2022, p. 14, Prop. 4.9]. Consequently, $\beta_0^{K,L}$ counts the connected components of K that remain disconnected in L , and the non-zero eigenvalues of $\Delta_0^{K,L}$ are related to effective resistances in L [Mémoli et al., 2022, p. 14, Thm. 4.10]. In the transaction-network context, a large value of $\beta_0^{K,L}$ indicates account clusters that are isolated at filtration level K but become connected only when the threshold is lowered to L — a pattern consistent with the compartmentalisation strategy characteristic of layering cascades (Section 2.3.3).

The persistent Hodge decomposition dual to Theorem 4.2.10 reads

$$C_q(K) = \ker \Delta_q^{K,L} \oplus \operatorname{im}(\partial_{q+1}^{K,L}) \oplus \operatorname{im}((\partial_q^K)^*), \quad (4.11)$$

extending the classical Hodge decomposition of Section 3.5.4 to the persistent setting [Liu et al., 2024, p. 12, Thm. 3.10]. We verify that the three summands are pairwise orthogonal:

- (i) $\ker \Delta_q^{K,L} \perp \operatorname{im}(\partial_{q+1}^{K,L})$: Every $c \in \ker \Delta_q^{K,L}$ satisfies $(\partial_{q+1}^{K,L})^* c = 0$ by (4.8). For any $\sigma \in C_{q+1}^{K,L}$, the definition of the adjoint yields:

$$\langle c, \partial_{q+1}^{K,L} \sigma \rangle = \langle (\partial_{q+1}^{K,L})^* c, \sigma \rangle = \langle 0, \sigma \rangle = 0.$$

- (ii) $\ker \Delta_q^{K,L} \perp \operatorname{im}((\partial_q^K)^*)$: Every $c \in \ker \Delta_q^{K,L}$ satisfies $\partial_q^K c = 0$ by (4.8). For any $u \in C_{q-1}(K)$, an element of the image is $(\partial_q^K)^* u$, and we have:

$$\langle c, (\partial_q^K)^* u \rangle = \langle \partial_q^K c, u \rangle = \langle 0, u \rangle = 0.$$

- (iii) $\operatorname{im}(\partial_{q+1}^{K,L}) \perp \operatorname{im}((\partial_q^K)^*)$: Let $\sigma \in C_{q+1}^{K,L}$ and $u \in C_{q-1}(K)$. Utilizing the boundary-of-a-boundary identity $\partial_q^K \circ \partial_{q+1}^{K,L} = 0$ (Definition 3.4.7, applied to $\partial_{q+1}^{K,L} \subset \partial_{q+1}^L$), we obtain:

$$\langle \partial_{q+1}^{K,L} \sigma, (\partial_q^K)^* u \rangle = \langle \partial_q^K (\partial_{q+1}^{K,L} \sigma), u \rangle = \langle 0, u \rangle = 0.$$

The three summands are thus pairwise orthogonal and span the entirety of $C_q(K)$ by a straightforward dimension count: their respective dimensions are $\beta_q^{K,L}$, $\operatorname{rank}(\partial_{q+1}^{K,L})$, and $\operatorname{rank}(\partial_q^K)$, yielding:

$$\beta_q^{K,L} + \operatorname{rank}(\partial_{q+1}^{K,L}) + \operatorname{rank}(\partial_q^K) = \dim \ker \partial_q^K + \operatorname{rank}(\partial_q^K) = \dim C_q(K),$$

where the first equality applies the Betti–nullity correspondence (4.7) and the second follows directly from Corollary 3.4.9 applied to ∂_q^K .

4.3 Schur-Complement Representation and Algorithms

The Schur-complement representation and the algorithmic constructions in this section are based on [Mémoli et al. \[2022, Secs. 3.1, 4.1–4.2\]](#). The definition of the persistent Laplacian and of the restricted boundary operator follows [Botnan and Dong \[2025, Def. 3.3\]](#), in which the up-persistent Laplacian is defined using the Hilbert-space adjoint of the restricted boundary operator.

For a symmetric block matrix

$$M = \begin{pmatrix} A & B \\ B^\top & D \end{pmatrix},$$

with D positive semidefinite, we use the *generalized Schur complement* of D in M , defined by

$$M/D := A - BD^\dagger B^\top.$$

In network theory, the operation that maps a graph Laplacian to the Schur complement with respect to a principal block associated with eliminated nodes is known as *Kron reduction* [[Dörfler and Bullo, 2013](#), p. 152, Eq. (4)]. In the notation of this thesis, Theorem [4.3.3](#) therefore shows that the up-persistent Laplacian $\Delta_{q,+}^{K,L}$ can be viewed as a Kron reduction of $\Delta_{q,+}^L$ onto the q -simplices of K .

The column-reduction construction in [Mémoli et al. \[2022, Lem. 3.4, Alg. 1\]](#) yields a matrix Z whose columns form a basis of the restricted chain group $C_{q+1}^{K,L} \subseteq C_{q+1}(L)$. Combining this construction with the Hilbert-space definition of the adjoint used for persistent Laplacians [[Botnan and Dong, 2025, Def. 3.3](#)], the orthogonal projection formula $P_{C_{q+1}^{K,L}} = Z(Z^\top Z)^{-1}Z^\top$, and the Moore–Penrose projection $P_{\ker B_{22}} = I - B_{22}^\dagger B_{22}$, gives the following ambient-space representation of the adjoint in Lemma [4.3.1](#):

$$(\partial_{q+1}^{K,L})^* = Z(Z^\top Z)^{-1}(\partial_{q+1}^{K,L})^\top.$$

Equivalently, its block form is

$$(\partial_{q+1}^{K,L})^* f = \begin{pmatrix} B_{11}^\top f \\ (I - B_{22}^\dagger B_{22})B_{12}^\top f \end{pmatrix}.$$

The Moore–Penrose identity $M^\dagger = M^\top (MM^\top)^\dagger$ (Lemma [4.3.2](#)) is a standard linear-algebra fact and is proved explicitly here for completeness.

4.3.1 Block Structure of the Boundary Matrix

Fix an inclusion $K \subseteq L$ and degree $q \geq 0$. Order the $(q+1)$ -simplices of L so that those of K come first, and order the q -simplices of L so that those of K come first. In these ordered bases the boundary matrix ∂_{q+1}^L takes the block form

$$\partial_{q+1}^L = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}, \tag{4.12}$$

where:

- $B_{11} \in \mathbb{R}^{n_q^K \times n_{q+1}^K}$: boundary of $(q+1)$ -simplices of K into q -simplices of K ; this is the matrix representation of ∂_{q+1}^K .
- $B_{12} \in \mathbb{R}^{n_q^K \times (n_{q+1}^L - n_{q+1}^K)}$: boundary of new $(q+1)$ -simplices in $L \setminus K$ into q -simplices of K .
- $B_{22} \in \mathbb{R}^{(n_q^L - n_q^K) \times (n_{q+1}^L - n_{q+1}^K)}$: boundary of new $(q+1)$ -simplices into new q -simplices.
- $0 \in \mathbb{R}^{(n_q^L - n_q^K) \times n_{q+1}^K}$: vanishes because no $(q+1)$ -simplex of K has a boundary component in $C_q(L) \setminus C_q(K)$; by Definition 3.6.6, the boundary of any simplex of K consists entirely of faces of K .

4.3.2 The Schur-Complement Formula

Before stating the main theorem, we record the form of the adjoint of the restricted boundary operator, which is used in the proof.

Lemma 4.3.1 (Matrix representation of the restricted adjoint). *Assume all simplicial complexes are unweighted ($w \equiv 1$). In the block notation of (4.12), the adjoint $(\partial_{q+1}^{K,L})^*$ satisfies*

$$(\partial_{q+1}^{K,L})^* f = \begin{pmatrix} B_{11}^\top f \\ (\text{Id} - B_{22}^\dagger B_{22}) B_{12}^\top f \end{pmatrix} \quad \text{for all } f \in C_q(K).$$

Proof. By definition, $(\partial_{q+1}^{K,L})^* f$ is the unique element of $C_{q+1}^{K,L} = C_{q+1}(K) \oplus \ker B_{22}$ satisfying

$$\langle (\partial_{q+1}^{K,L})^* f, c \rangle = \langle f, \partial_{q+1}^{K,L} c \rangle$$

for all $c = (c_1, c_2)^\top$ with $c_1 \in C_{q+1}(K)$ and $c_2 \in \ker B_{22}$. Given the block form of $\partial_{q+1}^{K,L}$, we have

$$\langle f, \partial_{q+1}^{K,L} c \rangle = \langle B_{11}^\top f, c_1 \rangle + \langle B_{12}^\top f, c_2 \rangle.$$

The first component is identified as $c_1^* = B_{11}^\top f$. For the second component, c_2 must lie in $\ker B_{22}$, so we project $B_{12}^\top f$ onto this kernel. Using the property of the Moore–Penrose pseudoinverse, where the orthogonal projection onto $\ker B_{22}$ is $P_{\ker B_{22}} = I - B_{22}^\dagger B_{22}$, we obtain

$$(\partial_{q+1}^{K,L})^* f = \begin{pmatrix} B_{11}^\top f \\ (I - B_{22}^\dagger B_{22}) B_{12}^\top f \end{pmatrix}. \quad (4.13)$$

This confirms the adjoint representation used in the derivation of the Schur complement. $\square$

Lemma 4.3.2 (Moore–Penrose identity). *For any real matrix M , we use the identity*

$$M^\dagger = M^\top (MM^\top)^\dagger.$$

Proof. Let $A := MM^\top$ and $X := M^\top A^\dagger$. Since A is symmetric, $A^\dagger$ is symmetric. One checks directly that X satisfies the Moore-Penrose conditions for M : $MXM = M$, $XM X = X$, $(MX)^\top = MX$, and $(XM)^\top = XM$. Hence $X = M^\dagger$. $\square$

Denote by $C_{11} \in \mathbb{R}^{n_{q-1}^K \times n_q^K}$ the matrix of ∂_q^K in the ordered simplex bases of $C_q(K)$ and $C_{q-1}(K)$. Note that C_{11} is distinct from B_{11} : the latter represents $\partial_{q+1}^K : C_{q+1}(K) \rightarrow C_q(K)$, while C_{11} represents $\partial_q^K : C_q(K) \rightarrow C_{q-1}(K)$.

Theorem 4.3.3 (Schur-complement representation [Mémoli et al., 2022, Thm. 4.6]). *In the block notation of (4.12), the up-persistent Laplacian satisfies*

$$\Delta_{q,+}^{K,L} = B_{11}B_{11}^\top + B_{12}(\text{Id} - B_{22}^\dagger B_{22})B_{12}^\top = B_{11}B_{11}^\top + B_{12}B_{12}^\top - B_{12}B_{22}^\dagger B_{22}B_{12}^\top, \quad (4.14)$$

where $B_{22}^\dagger$ denotes the Moore-Penrose pseudoinverse of B_{22} .

Equivalently, $\Delta_{q,+}^{K,L}$ is the Schur complement of the $(L \setminus K)$ -principal block of $\Delta_{q,+}^L$. Consequently,

$$\Delta_q^{K,L} = \underbrace{B_{11}B_{11}^\top + B_{12}B_{12}^\top - B_{12}B_{22}^\dagger B_{22}B_{12}^\top}_{\Delta_{q,+}^{K,L}} + \underbrace{C_{11}^\top C_{11}}_{\Delta_{q,-}^K}. \quad (4.15)$$

The proof follows the block-matrix structure implicit in Mémoli et al. [2022, Thm. 4.6], specialized to the unweighted case $w \equiv 1$ and the simplex bases introduced in Section 4.3.2.

Proof. We derive (4.14) from Definition 4.2.4 using Lemma 4.3.1.

Step 1: Applying the adjoint. By Lemma 4.3.1, for any $f \in C_q(K)$,

$$(\partial_{q+1}^{K,L})^* f = (B_{11}^\top f, (\text{Id} - B_{22}^\dagger B_{22})B_{12}^\top f)^\top.$$

Step 2: Applying $\partial_{q+1}^{K,L}$. Setting $c_1 = B_{11}^\top f$ and $c_2 = (\text{Id} - B_{22}^\dagger B_{22})B_{12}^\top f$,

$$\begin{aligned} \Delta_{q,+}^{K,L} f &= \partial_{q+1}^{K,L} ((\partial_{q+1}^{K,L})^* f) = B_{11}c_1 + B_{12}c_2 \\ &= B_{11}B_{11}^\top f + B_{12}(\text{Id} - B_{22}^\dagger B_{22})B_{12}^\top f, \end{aligned} \quad (4.16)$$

which is the first equality in (4.14).

Step 3: Identification as Schur complement. Expanding (4.16) gives

$$\Delta_{q,+}^{K,L} f = B_{11}B_{11}^\top f + B_{12}B_{12}^\top f - B_{12}B_{22}^\dagger B_{22}B_{12}^\top f.$$

We verify that this coincides with the Schur complement of the $(L \setminus K)$ -principal block of $\Delta_{q,+}^L = \partial_{q+1}^L (\partial_{q+1}^L)^*$. From (4.12), the block decomposition of $\Delta_{q,+}^L$ is

$$[\Delta_{q,+}^L]_{KK} = B_{11}B_{11}^\top + B_{12}B_{12}^\top, \quad [\Delta_{q,+}^L]_{K,L \setminus K} = B_{12}B_{22}^\top, \quad [\Delta_{q,+}^L]_{L \setminus K, L \setminus K} = B_{22}B_{22}^\top.$$

The Schur complement of $B_{22}B_{22}^\top$ in $\Delta_{q,+}^L$ is

$$\begin{aligned} & [\Delta_{q,+}^L]_{KK} - [\Delta_{q,+}^L]_{K,L \setminus K} (B_{22}B_{22}^\top)^\dagger [\Delta_{q,+}^L]_{L \setminus K, K} \\ &= B_{11}B_{11}^\top + B_{12}B_{12}^\top - B_{12}B_{22}^\top (B_{22}B_{22}^\top)^\dagger B_{22}B_{12}^\top. \end{aligned}$$

By Lemma 4.3.2, we have the identity $B_{22}^\top (B_{22}B_{22}^\top)^\dagger = B_{22}^\dagger$. Substituting this into the Schur complement expansion yields $B_{22}^\top (B_{22}B_{22}^\top)^\dagger B_{22} = B_{22}^\dagger B_{22}$, and hence

$$B_{11}B_{11}^\top + B_{12}B_{12}^\top - B_{12}B_{22}^\dagger B_{22}B_{12}^\top = \Delta_{q,+}^{K,L},$$

confirming (4.14). The formula $\Delta_{q,-}^K = C_{11}^\top C_{11}$ is immediate from Definition 3.5.5. $\square$

4.3.3 Algorithm 1: Direct Construction

The following procedure computes $\Delta_q^{K,L}$ directly from the block decomposition (4.12) and the explicit adjoint representation of Lemma 4.3.1.

Input. A simplicial pair $K \subseteq L$ and a degree $q \geq 0$.

Output. The matrix of $\Delta_q^{K,L}$ in the simplex basis of $C_q(K)$.

Steps.

1. Enumerate the q -simplices of L and the $(q+1)$ -simplices of L , with all simplices of K ordered first in each list. Form the boundary matrix ∂_{q+1}^L in these ordered simplex bases and extract the blocks B_{11} , B_{12} , B_{22} as in (4.12).

2. Form the matrix

$$D_{q+1}^L := \partial_{q+1}^L ([n_q^L] \setminus [n_q^K]; \cdot),$$

i.e. the submatrix of ∂_{q+1}^L consisting of the rows indexed by q -simplices in $L \setminus K$. This matrix represents the boundary map $\partial_{q+1}^L : C_{q+1}(L) \rightarrow C_q(L \setminus K)$.

3. Perform column reduction on D_{q+1}^L to obtain

$$R_{q+1}^L = D_{q+1}^L Y,$$

where Y is an invertible change-of-basis matrix.

4. Let I be the index set of zero columns of R_{q+1}^L and set $Z := Y(:, I)$. Then the columns of Z form a basis of the restricted chain group $C_{q+1}^{K,L} \subseteq C_{q+1}(L)$ as in Mémoli et al. [2022, Lem. 3.4, Alg. 1].

5. Using Lemma 4.3.1, compute the matrix of the restricted adjoint

$$(\partial_{q+1}^{K,L})^* = Z(Z^\top Z)^{-1}(\partial_{q+1}^{K,L})^\top,$$

where $\partial_{q+1}^{K,L}$ is the restriction of ∂_{q+1}^L to $C_{q+1}^{K,L}$ in the basis Z .

6. Form the up-persistent part as

$$\Delta_{q,+}^{K,L} = \partial_{q+1}^{K,L}(\partial_{q+1}^{K,L})^*,$$

and the down-part as

$$\Delta_{q,-}^K = (\partial_q^K)^\top \partial_q^K,$$

where ∂_q^K is the boundary of K in degree q .

7. Return the full persistent Laplacian

$$\Delta_q^{K,L} = \Delta_{q,+}^{K,L} + \Delta_{q,-}^K.$$

Complexity. The dominant cost arises from assembling the boundary matrices of L and performing the column reduction on D_{q+1}^L . In the intended application, this direct construction is typically less efficient than Algorithm 2, since the latter reuses the precomputed up-Laplacian $\Delta_{q,+}^L$ across all choices of $K \subseteq L$.

4.3.4 Algorithm 2: Via the Schur complement

A simpler alternative computes $\Delta_q^{K,L}$ by first forming the up-persistent Laplacian via a Schur complement of the up-Laplacian of L and then adding the down-part of K .

Input. A simplicial pair $K \subseteq L$ and a degree $q \geq 0$.

Output. The matrix of $\Delta_q^{K,L}$ in the simplex basis of $C_q(K)$.

Steps.

1. Compute the up-Laplacian

$$\Delta_{q,+}^L = \partial_{q+1}^L(\partial_{q+1}^L)^\top$$

in the ordered basis where q -simplices of K appear first.

2. Let I_L^K denote the index set of q -simplices of K and $J_L^K := [n_q^L] \setminus I_L^K$ the index set of q -simplices in $L \setminus K$. Partition

$$\Delta_{q,+}^L = \begin{pmatrix} U_{KK} & U_{KJ} \\ U_{JK} & U_{JJ} \end{pmatrix}$$

according to I_L^K and J_L^K , and compute the Schur complement of the $(L \setminus K)$ -principal block:

$$\Delta_{q,+}^{K,L} := U_{KK} - U_{KJ}U_{JJ}^\dagger U_{JK}.$$

3. Compute the down-part of K ,

$$\Delta_{q,-}^K = (\partial_q^K)^\top \partial_q^K,$$

in the simplex basis of $C_q(K)$.

4. Return the full persistent Laplacian

$$\Delta_q^{K,L} = \Delta_{q,+}^{K,L} + \Delta_{q,-}^K.$$

Complexity. Algorithm 2 is preferred over Algorithm 1 in the AML pipeline of Chapter 5 because $\Delta_{q,+}^L$ is precomputed once per filtration step and reused across all choices of $K \subseteq L$, amortising the assembly cost.

Having derived two algorithms for computing $\Delta_q^{K,L}$, we now study how its spectrum behaves under perturbations of the filtration.

4.4 Spectral Properties and Stability

4.4.1 Eigenvalue Monotonicity

The key spectral consequence of the Schur-complement representation (Section 4.3) is that extending the ambient complex can only *increase* the up-persistent eigenvalues. The following lemma, which underpins all subsequent monotonicity statements, is a consequence of the Loewner order on positive semidefinite matrices.

Lemma 4.4.1 (PSD dominance under Schur complement [Mémoli et al., 2022, Lem. 4.2, Lem. 4.5]). *Let M be a symmetric positive semidefinite matrix partitioned as*

$$M = \begin{pmatrix} A & C \\ C^\top & D \end{pmatrix},$$

with D positive semidefinite, and let $M/D := A - CD^\dagger C^\top$ denote the Schur complement of D in M . Then $M/D \preceq A$ in the Loewner order, i.e.,

$$\langle (M/D)v, v \rangle \leq \langle Av, v \rangle \quad \text{for all } v. \quad (4.17)$$

Here, for symmetric matrices X, Y we write $X \preceq Y$ if and only if $Y - X$ is positive semidefinite; this is the Loewner partial order on Hermitian matrices (see, for example, Loewner-Halbordnung [Wikipedia contributors, b]).

Proof. Since $M \succeq 0$, for every vector $u = (v, w)^\top$,

$$0 \leq \langle Mu, u \rangle = \langle Av, v \rangle + 2\langle C^\top v, w \rangle + \langle Dw, w \rangle.$$

Minimising this quadratic form over w yields the minimum-norm solution $w^* = -D^\dagger C^\top v$. Since $M \succeq 0$, the block factorisation $M = EE^\top$ used in the proof of Lemma 4.2 in Appendix A of [Mémoli et al. \[2022\]](#) implies $\text{im}(C^\top) \subseteq \text{im}(D)$, from which it follows that $DD^\dagger C^\top v = C^\top v$. Substituting w^* back into the expansion simplifies the cross-terms and yields

$$0 \leq \langle Av, v \rangle - \langle CD^\dagger C^\top v, v \rangle = \langle (M/D)v, v \rangle, \quad (4.18)$$

which simultaneously proves $M/D \succeq 0$ and $M/D \preceq A$. $\square$

From the Loewner order $M/D \preceq A$, one immediately deduces a pointwise inequality on eigenvalues via the Courant–Fischer Min–Max theorem (also known as the variational, or minimum–maximum, principle; see, for example, *Satz von Courant-Fischer* [[Wikipedia contributors, a](#)]). For a symmetric (or Hermitian) matrix X with eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$, this theorem states that its k -th eigenvalue can be written as

$$\lambda_k(X) = \min_{\dim V=k} \max_{v \in V \setminus \{0\}} \frac{\langle Xv, v \rangle}{\|v\|^2},$$

which implies that, since both M/D and A are symmetric positive semidefinite matrices of size $n \times n$,

$$\lambda_k(M/D) = \min_{\dim V=k} \max_{v \in V} \frac{\langle (M/D)v, v \rangle}{\|v\|^2} \leq \min_{\dim V=k} \max_{v \in V} \frac{\langle Av, v \rangle}{\|v\|^2} = \lambda_k(A) \quad (4.19)$$

for each $k = 1, \dots, n$, because the Rayleigh quotient of M/D does not exceed that of A for any given vector v . The following lemma records the quotient property of Schur complements that is needed in the proof of the main monotonicity theorem.

Lemma 4.4.2 (Quotient property of Schur complements [[Mémoli et al., 2022](#), Lem. 4.4]). *Let M be a symmetric positive semidefinite matrix partitioned conformally as*

$$M = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{12}^\top & A_{22} & A_{23} \\ A_{13}^\top & A_{23}^\top & A_{33} \end{pmatrix}.$$

Let A_{22-33} denote the compound lower-right block matrix $\begin{pmatrix} A_{22} & A_{23} \\ A_{23}^\top & A_{33} \end{pmatrix}$. Then the Schur complement of A_{22-33} in M equals the Schur complement of (A_{22-33}/A_{33}) in the Schur complement of A_{33} in M :

$$M/A_{22-33} = (M/A_{33})/(A_{22-33}/A_{33}). \quad (4.20)$$

In particular, for a simplicial pair $K \subseteq K' \subseteq L$, successive Schur complementation of $C_q(L \setminus K')$ and then $C_q(K' \setminus K)$ yields the same result as simultaneously complementing $C_q(L \setminus K)$.

Proof. The identity (4.20) is a standard property of Schur complements of positive semidefinite matrices; the formal algebraic proof is given in [Mémoli et al., 2022, Lem. 4.4]. The simplicial interpretation follows directly from the block structure of $\Delta_{q,+}^L$ with respect to the orthogonal splitting $C_q(K_{t_1}) \oplus C_q(K_{t_2} \setminus K_{t_1}) \oplus C_q(K_{t_3} \setminus K_{t_2})$ via Theorem 4.3.3. $\square$

Applying Lemma 4.4.1 and Lemma 4.4.2 to the block decomposition of $\Delta_{q,+}^L$ now yields the central monotonicity result.

Theorem 4.4.3 (Monotonicity of up-persistent eigenvalues [Mémoli et al., 2022, Thm. 5.2]). *Let $\mathcal{K} = \{K_t\}_{t \in T}$ be a simplicial filtration in the sense of Definition 3.6.1 and fix $q \geq 0$. For any $t_1 \leq t_2 \leq t_3$ in T and each $k = 1, \dots, n_q^{K_{t_1}}$,*

$$\lambda_{q,+,k}(K_{t_1}, K_{t_2}) \leq \lambda_{q,+,k}(K_{t_1}, K_{t_3}) \quad \text{and} \quad \lambda_{q,+,k}(K_{t_2}, K_{t_3}) \leq \lambda_{q,+,k}(K_{t_1}, K_{t_3}). \quad (4.21)$$

Proof. We prove the first inequality; the second follows from an independent interlacing argument detailed at the end of Step 3.

Step 1: Block decomposition. Since $K_{t_1} \subseteq K_{t_3}$, the full up-Laplacian $\Delta_{q,+}^{K_{t_3}}$ admits a block decomposition

$$\Delta_{q,+}^{K_{t_3}} = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix}$$

with respect to the splitting $C_q(K_{t_1}) \oplus C_q(K_{t_3} \setminus K_{t_1})$. By Theorem 4.3.3, the persistent up-Laplacian $\Delta_{q,+}^{K_{t_1}, K_{t_3}}$ equals the Schur complement of A_{22} in $\Delta_{q,+}^{K_{t_3}}$:

$$\Delta_{q,+}^{K_{t_1}, K_{t_3}} = A_{11} - A_{12} A_{22}^\dagger A_{12}^\top. \quad (4.22)$$

Step 2: Quotient property. Since $K_{t_1} \subseteq K_{t_2} \subseteq K_{t_3}$, we further partition the complement space as

$$C_q(K_{t_3} \setminus K_{t_1}) = C_q(K_{t_2} \setminus K_{t_1}) \oplus C_q(K_{t_3} \setminus K_{t_2}),$$

giving the refined block decomposition

$$\Delta_{q,+}^{K_{t_3}} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{12}^\top & A_{22} & A_{23} \\ A_{13}^\top & A_{23}^\top & A_{33} \end{pmatrix}$$

where A_{11} is indexed by $C_q(K_{t_1})$, the middle block by $C_q(K_{t_2} \setminus K_{t_1})$, and the lower block by $C_q(K_{t_3} \setminus K_{t_2})$. By Theorem 4.3.3, the persistent up-Laplacian $\Delta_{q,+}^{K_{t_1}, K_{t_2}}$ is the Schur complement of the lower-right block A_{22-33} in $\Delta_{q,+}^{K_{t_3}}$. By the quotient property (Lemma 4.4.2), this step avoids loss of generality and equals the sequential Schur complementation of A_{33} within $\Delta_{q,+}^{K_{t_1}, K_{t_3}}$. Hence, Lemma 4.4.1 directly implies the Loewner ordering

$$\Delta_{q,+}^{K_{t_1}, K_{t_2}} \preceq \Delta_{q,+}^{K_{t_1}, K_{t_3}}. \quad (4.23)$$

Step 3: Eigenvalue comparison. The Loewner order (4.23) combined with the Courant–Fischer inequality (4.19) yields, for each k ,

$$\lambda_{q,+,k}(K_{t_1}, K_{t_2}) \leq \lambda_{q,+,k}(K_{t_1}, K_{t_3}),$$

confirming the first inequality of (4.21).

Second inequality of (4.21). We now prove $\lambda_{q,+,k}(K_{t_2}, K_{t_3}) \leq \lambda_{q,+,k}(K_{t_1}, K_{t_3})$. Utilizing the index notation $I_{ts} := [n_q^{K_t}] \setminus [n_q^{K_s}]$ for $s \leq t$, Theorem 4.3.3 states that

$$\Delta_{q,+}^{K_{t_2}, K_{t_3}} = \Delta_{q,+}^{K_{t_3}} / I_{t_3 t_2}, \quad \Delta_{q,+}^{K_{t_1}, K_{t_3}} = \Delta_{q,+}^{K_{t_3}} / I_{t_3 t_1}.$$

Since $K_{t_1} \subseteq K_{t_2}$, we have $I_{t_3 t_2} \subseteq I_{t_3 t_1}$. Thus, both operators represent Schur complements of proper principal submatrices within $\Delta_{q,+}^{K_{t_3}}$ [Mémoli et al., 2022, Def. 4.1]. Invoking the formal Quotient Formula (Lemma 4.4.2), $\Delta_{q,+}^{K_{t_3}} / I_{t_3 t_1}$ can be viewed as a further Schur complement of a proper principal submatrix inside $\Delta_{q,+}^{K_{t_3}} / I_{t_3 t_2}$. The eigenvalue interlacing property for Schur complements of submatrices [Mémoli et al., 2022, Lem. 4.5] then establishes that

$$\lambda_k(\Delta_{q,+}^{K_{t_3}} / I_{t_3 t_1}) \leq \lambda_k(\Delta_{q,+}^{K_{t_3}} / I_{t_3 t_2}),$$

which immediately translates to

$$\lambda_{q,+,k}(K_{t_2}, K_{t_3}) \leq \lambda_{q,+,k}(K_{t_1}, K_{t_3})$$

for each valid index $k = 1, \dots, n_q^{K_{t_1}}$ [Mémoli et al., 2022, Thm. 5.2]. $\square$

Remark 4.4.4. The monotonicity in (4.21) concerns the *up*-persistent Laplacian $\Delta_{q,+}^{K,L}$ only. The down-Laplacian $\Delta_{q,-}^K$ depends solely on K [Mémoli et al., 2022, Eq. (3)] and remains invariant as the ambient complex grows; the full persistent Laplacian $\Delta_q^{K,L} = \Delta_{q,+}^{K,L} + \Delta_{q,-}^K$ therefore inherits monotonicity in its up-part exclusively.

In the AML context, Theorem 4.4.3 guarantees that the spectral signal cannot be weakened by adding later transactions to the ambient complex: if the eigenvalue $\lambda_{q,+,k}(K_{t_1}, K_{t_2}) > 0$ is detected for a fraud-ring window $[t_1, t_2]$, then the corresponding eigenvalue at any larger window $[t_1, t_3] \supseteq [t_1, t_2]$ is at least as large. This monotone non-decreasing behaviour, established for up-persistent eigenvalues in Section 5.3 of Mémoli et al. [2022], shows that the spectral signal cannot be diluted by later, unrelated transactions entering the ambient complex.

4.4.2 Algebraic Stability

Theorem 4.4.3 controls individual eigenvalues within a fixed filtration. A stronger result, due to Liu et al. [2024], bounds the entire collection of persistent Laplacians under perturbation of the filtration itself.

Definition 4.4.5 (Persistence differential graded inner product space (DGI)). A *persistence DGI space* $S: (\mathbb{R}, \leq) \rightarrow \mathbf{DGI}$ is a functor from the poset $(\mathbb{R}, \leq)$ to the category of differential graded real inner product spaces and isometric chain maps. In the present setting, S assigns to each $t \in \mathbb{R}$ the chain complex $(C_\bullet(K_t), \partial_\bullet, \langle \cdot, \cdot \rangle)$ and to each $s \leq t$ the inclusion-induced chain map $K_s \hookrightarrow K_t$.

For a persistence DGI space S , denote by $\mathbf{LS}$ the associated persistence Laplacian tree constructed in Liu et al. [2024, Sec. 4].

Theorem 4.4.6 (Algebraic stability [Liu et al., 2024, Thm. 5.1, Cor. 5.2]). *Let $S, T: (\mathbb{R}, \leq) \rightarrow \mathbf{DGI}$ be two persistence DGI spaces over the same underlying vertex set. Then:*

1. *S and T are ε -interleaved if and only if $\mathbf{LS}$ and $\mathbf{LT}$ are ε -interleaved. Consequently,*

$$d_I(\mathbf{LS}, \mathbf{LT}) = d_I(S, T). \quad (4.24)$$

2. [Liu et al., 2024, Thm. 5.3] *For two non-decreasing functions $f, g: K \rightarrow \mathbb{R}$ defined on all simplices of a simplicial complex K , Liu et al. [2024] formulate the result for the sublevel-set filtration $K_a^f := f^{-1}((-\infty, a])$; here we use the equivalent descending-threshold convention of Remark 4.7.4,*

$$K_t^f := \{\sigma \in K : f(\sigma) \geq t\},$$

(i.e., $K_s^f \supseteq K_t^f$ for $s \leq t$, opposite to the ascending convention of Definition 3.6.1),

$$d_I(\mathbf{L}_f, \mathbf{L}_g) \leq \|f - g\|_\infty. \quad (4.25)$$

Proof sketch. We sketch the proof of part (1); part (2) follows directly by combining (1) with the standard stability of filtrations under $\|\cdot\|_\infty$ perturbations: if $\|f - g\|_\infty \leq \varepsilon$, then $K_t^f \subseteq K_{t-\varepsilon}^g$ for all t , implying that S_f and S_g are ε -interleaved [Liu et al., 2024, Thm. 5.3], and part (1) then yields $d_I(\mathbf{L}_f, \mathbf{L}_g) \leq \varepsilon = \|f - g\|_\infty$.

($\Rightarrow$) Interleaving of filtrations implies interleaving of Laplacian trees. By Liu et al. [2024, Thm. 5.1], the functor $\mathbf{L}: \mathbf{DGI}_\mathbb{R} \rightarrow \mathbf{Lap}_\mathbb{R}$ constitutes an isometry with respect to the interleaving distance. Concretely, if S and T are ε -interleaved via morphisms $\phi: S \rightarrow \varepsilon T$ and $\psi: T \rightarrow \varepsilon S$ in $\mathbf{DGI}_\mathbb{R}$ [Liu et al., 2024, Def. 2.1], applying the functor $\mathbf{L}$ to the two interleaving diagrams produces valid morphisms $\mathbf{L}\phi: \mathbf{LS} \rightarrow \varepsilon \mathbf{LT}$ and $\mathbf{L}\psi: \mathbf{LT} \rightarrow \varepsilon \mathbf{LS}$ in $\mathbf{Lap}_\mathbb{R}$ satisfying the interleaving compatibility conditions, so $\mathbf{LS}$ and $\mathbf{LT}$ are ε -interleaved.

($\Leftarrow$) Interleaving of Laplacian trees implies interleaving of filtrations. The converse direction is established in Liu et al. [2024, Thm. 5.1] via the recovery functor $P: \mathbf{Lap}_\mathbb{R} \rightarrow \mathbf{DGI}_\mathbb{R}$ defined by $P(I)_a = I_a$ for each $a \in \mathbb{R}$: if $\mathbf{LS}$ and $\mathbf{LT}$ are ε -interleaved in $\mathbf{Lap}_\mathbb{R}$, applying P fully recovers an ε -interleaving between $PLS = S$ and $PLT = T$ in $\mathbf{DGI}_\mathbb{R}$. $\square$

Theorem 4.4.6 is strictly stronger than Theorem 4.4.3: whereas Theorem 4.4.3 bounds individual eigenvalues within a fixed filtration, Theorem 4.4.6 controls the entire Laplacian tree under an ε -perturbation of the underlying filtration. The earlier scalar stability result of Mémoli et al. [2022, Thm. 5.9], which bounds the interleaving distance between the eigenvalue functions $I \mapsto \lambda_{q,+k}(I)$ by the filtration interleaving distance $d_{\text{VI}}(\mathcal{K}, \mathcal{L})$ via Lemma 5.5 of Mémoli et al. [2022], is an independent spectral stability result at the scalar level and is complementary to the categorical stability of (4.24).

For AML transaction networks, inequality (4.25) implies that two filtrations differing by at most ε in L^∞ -norm produce Laplacian trees whose interleaving distance is at most ε . Small day-to-day data-quality differences therefore cause only small spectral perturbations of the resulting Laplacian tree. The empirical sensitivity of the detection pipeline to the hyperparameters W (window length) and T (number of volume steps) is discussed in Section 5.6.4. The bound (4.25) implies that perturbations $|\delta w| < \min_e w_{\text{sym}}(e)$ that do not alter which edges satisfy $w_{\text{sym}}(e) \geq \theta$ for any threshold θ in the sequence $\theta_0 > \dots > \theta_T$ (Remark 4.7.4) leave the Laplacian-tree feature vector essentially unchanged, which is particularly relevant for the smurfing scenario of Section 2.3.2 where transactions are kept deliberately close to the reporting threshold.

The algebraic stability established above applies globally to the Laplacian tree. We now derive a pointwise geometric bound for the case $q = 0$, the persistent Cheeger inequality.

4.4.3 Persistent Cheeger Inequality

Recall from Section 3.3 that the classical discrete Cheeger inequality relates the Fiedler value $\lambda_1(L_{\text{sym}})$ of the symmetrised graph Laplacian to the edge-expansion of the graph (Definition 3.3.10, [Mémoli et al., 2022, Eq. (11)]). The following definition and theorem establish a persistent analogue for graph pairs $K \hookrightarrow L$ (i.e., $q = 0$), using the identification of $\Delta_0^{K,L}$ with a Kron-reduced Laplacian.

Here $R^L(A, B)$ denotes the effective resistance between vertex sets A and B in the graph L , defined via the Moore–Penrose pseudoinverse of its Laplacian [Mémoli et al., 2022, Sec. 4.3].

Definition 4.4.7 (Persistent Cheeger constant [Mémoli et al., 2022, Def. 4.17]). Let $K \hookrightarrow L$ be a connected weighted graph pair with vertex set $V(K)$. For a subset $A \subseteq V(K)$, write $C^L(A, V(K) \setminus A) := (R^L(A, V(K) \setminus A))^{-1}$ for the effective conductance between A and its complement, computed with respect to the edge weights of L . The *persistent Cheeger constant* of the pair (K, L) is

$$h^{K,L} = \min_{\substack{A \subseteq V(K) \\ 0 < |A| \leq \frac{1}{2}|V(K)|}} \frac{C^L(A, V(K) \setminus A)}{|A|}. \quad (4.26)$$

When $K = L$, this reduces to the classical Cheeger constant $h(K)$.

Theorem 4.4.8 (Kron reduction preserves effective resistances [Mémoli et al., 2022, Thm. 4.15]). *Let $K \hookrightarrow L$ be a connected weighted graph pair and let $\tilde{K}$ denote the Kron-reduced graph on $V(K)$, so that $\Delta_0^{K,L} = L_{\tilde{K}}$ (Prop. 4.9 of [Mémoli et al., 2022]). Then for all disjoint $A, B \subseteq V(K)$,*

$$R^{\tilde{K}}(A, B) = R^L(A, B). \quad (4.27)$$

The proof is made in [Mémoli et al. \[2022, p. 16\]](#).

The persistent Cheeger constant is monotone non-decreasing in the ambient complex: for any $K \hookrightarrow L_1 \hookrightarrow L_2$,

$$h^{K,L_1} \leq h^{K,L_2}, \quad (4.28)$$

since adding edges to L can only decrease effective resistances and hence increase conductances [[Mémoli et al., 2022, Prop. 4.18](#)].

Theorem 4.4.9 (Persistent Cheeger inequality [[Mémoli et al., 2022, Thm. 4.20](#)]). *Let $K \hookrightarrow L$ be a connected weighted graph pair in the sense of Definition 4.4.7. Then*

$$\frac{(h^{K,L})^2}{2 d_{\max}^{K,L}} \leq \lambda_{+,2}^{(0)}(K, L) \leq 2 h^{K,L}, \quad (4.29)$$

for the maximum weighted degree in L over vertices of K $d_{\max}^{K,L} := \max_{v \in V(K)} d_v(L)$, and the second-smallest eigenvalue of $\Delta_0^{K,L}$ $\lambda_{+,2}^{(0)}(K, L)$.

The proof is made in [Mémoli et al. \[2022, p. 18\]](#).

Remark 4.4.10 (Higher-dimensional Cheeger inequality). Theorem 4.4.9 is currently proven only for $q = 0$ (graph pairs). For higher-dimensional *non-branching* simplicial pairs, [Botnan and Dong \[2025, Thm. 4.10\]](#) establish a two-sided Cheeger-type inequality for the up-persistent Laplacian, under the assumptions that L is *unweighted* and q -non-branching, that the boundary matrix B_{q+1}^L has full column rank, and that $B_{q+1}^{L,K} \neq 0$:

$$\min_{j \leq h+r} \frac{A_b(\mathcal{P}_j)}{V(\mathcal{P}_j)} \leq \lambda_{\min}^{K,L} \leq \min_{j \leq h+r} \frac{A(\mathcal{P}_j)}{V(\mathcal{P}_j)}, \quad (4.30)$$

where $\mathcal{P}_j$, $1 \leq j \leq h+r$, are the connected $(q+1)$ -dimensional polyhedra formed by the $(q+1)$ -simplices of L corresponding to the columns of $B_{q+1}^{L,K}$ [[Botnan and Dong, 2025, Sec. 4.2](#)]. Here $A(\mathcal{P}_j)$ is the number of q -simplices forming the boundary of $\mathcal{P}_j$, $V(\mathcal{P}_j)$ is the number of $(q+1)$ -simplices in $\mathcal{P}_j$, and $A_b(\mathcal{P}_j) := A(\mathcal{P}_j) - \sum_{i \neq j} A(\mathcal{P}_i \cap \mathcal{P}_j)$ denotes the number of boundary q -simplices of $\mathcal{P}_j$ not shared with any other polyhedron [[Botnan and Dong, 2025, Sec. 4.2](#)] (we write $\hat{A}$ for this quantity in place of Botnan's A_b notation). A full higher-dimensional analogue of Definition 4.4.7 for general simplicial pairs (i.e., beyond the q -non-branching case) remains an open problem; see the discussion following [Botnan and Dong \[2025, Thm. 4.10\]](#).

Inequality (4.29) links the spectral feature $\lambda_{+,2}^{(0)}(K_t, K_{t+1})$ directly to the edge expansion of the transaction graph at filtration scale t : a small persistent spectral gap certifies a bottleneck structure characteristic of layering cascades (Section 2.3.3), while a large gap indicates a well-connected, ring-shaped topology whose structure is robust to small perturbations. The monotonicity (4.28) of $h^{K,L}$ further implies that adding more transactions at the next filtration step can only sharpen this spectral signal. In the higher-dimensional setting of transaction triangles and three-way collusion rings (Section 2.3.1),

the geometric interpretation of (4.30) is particularly transparent for unweighted, q -non-branching transaction complexes: the minimal non-zero eigenvalue is bounded between two isoperimetric ratios, from below by the ratio of exclusive boundary q -simplices to volume and from above by the ratio of all boundary q -simplices to volume [Botnan and Dong, 2025, Thm. 4.10].

4.5 Persistent Laplacians on Non-Branching Complexes

4.5.1 Motivation: Transaction Chains

Layering cascades (Section 2.3.3) often take the form of linear transfer chains or simple directed cycles. After symmetrisation via the procedure of Section 3.2, these produce non-branching one-dimensional simplicial complexes, for which the persistent Laplacian admits a closed-form spectral analysis based on the Schur-complement characterisation of Theorem 4.3.3.

Definition 4.5.1 (Non-branching complex). A one-dimensional simplicial complex L is called *non-branching* if every vertex of L is the face of at most two edges, i.e., $d(v) \leq 2$ for all $v \in L$, or equivalently each *column* of the boundary matrix $\partial_1^L \in \mathbb{R}^{n_1^L \times n_0^L}$ (rows indexed by edges, columns by vertices) contains at most two non-zero entries.

A one-dimensional simplicial complex is non-branching in the sense of Botnan and Dong [2025, Def. 3.1] precisely when it is 0-non-branching: every vertex is a face of at most two edges. Equivalently, its underlying graph has maximum degree at most two, so every connected component is an isolated vertex, a path, or a cycle.

In the AML setting, a path P_k on k vertices represents a $(k-1)$ -step layering cascade (Section 2.3.3), while a simple cycle C_n represents a closed transaction ring (Section 2.3.1).

Throughout this section we write $\Delta_{1,+}^{P_k, P_n}$ for $\Delta_{1,+}^{K, L}$ with $K = P_k$, $L = P_n$ in the sense of Definition 3.6.6.

4.5.2 Spectrum of the Path Gram Matrix

Before computing persistent spectra, we recall the eigenvalues of the classical Gram matrix

$$T_n := (\partial_1^{P_n})^\top \partial_1^{P_n}$$

for the path graph P_n ; these follow from the spectrum of the path Laplacian [Spielman, 2025, Thm. 5.6.1].

Lemma 4.5.2 (Eigenvalues of the path Gram matrix). *Let P_n be the path graph on n vertices, with signed incidence matrix $\partial_1^{P_n} \in \mathbb{R}^{(n-1) \times n}$. The Gram matrix*

$$T_n := (\partial_1^{P_n})^\top \partial_1^{P_n} \in \mathbb{R}^{n \times n}$$

is the symmetric tridiagonal matrix

$$(T_n)_{ij} = \begin{cases} 1 & \text{if } i = j \in \{1, n\}, \\ 2 & \text{if } i = j \in \{2, \dots, n-1\}, \\ -1 & \text{if } |i - j| = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (4.31)$$

Its eigenvalues are

$$\mu_j(T_n) = 4 \sin^2\left(\frac{j\pi}{2n}\right), \quad j = 0, 1, \dots, n-1. \quad (4.32)$$

In particular, $\mu_0 = 0$ and $\mu_j > 0$ for $j \geq 1$.

Proof. The tridiagonal form (4.31) follows by a direct computation from the signed incidence matrix of the path graph: each interior vertex has combinatorial degree 2, each endpoint has degree 1, and each edge contributes -1 to the two adjacent off-diagonal entries.

For an unweighted graph, the standard graph Laplacian can be written as $L = \partial_1^\top \partial_1$, so T_n coincides with the Laplacian of P_n . By Spielman [2025, Thm. 5.6.1], the Laplacian of P_n has eigenvalues

$$2\left(1 - \cos\left(\frac{j\pi}{n}\right)\right), \quad j = 0, \dots, n-1.$$

Using the elementary trigonometric identity $2(1 - \cos x) = 4 \sin^2(x/2)$ with $x = \frac{j\pi}{n}$, we obtain

$$2\left(1 - \cos\left(\frac{j\pi}{n}\right)\right) = 4 \sin^2\left(\frac{j\pi}{2n}\right), \quad j = 0, \dots, n-1,$$

which yields (4.32). $\square$

Remark 4.5.3 (Role of Lemma 4.5.2). Lemma 4.5.2 gives the closed-form spectrum of the vertex Laplacian $(\partial_1^{P_n})^\top \partial_1^{P_n}$. In Theorem 4.5.4, the complementary block D in (4.37) is instead the edge Gram matrix $\partial_1^{P_{m+1}}(\partial_1^{P_{m+1}})^\top$, which shares the same non-zero eigenvalues but is invertible. Its explicit inverse yields $(D^{-1})_{1,1}$ in closed form, which in turn gives the explicit tridiagonal representation of $\Delta_{1,+}^{P_k, P_n}$.

4.5.3 Explicit Spectra for Path Inclusions

The explicit spectral analysis of path inclusions serves as a calibration tool for AML: a layering cascade of depth $n - k$, modelled by the inclusion $P_k \subseteq P_n$, induces a persistent up-Laplacian whose smallest eigenvalue is governed by the boundary equation (4.36), as quantified in Theorem 4.5.4. For fixed retained length k , this smallest eigenvalue empirically decreases in the extension length $m = n - k$ (though we do not establish strict monotonicity for all m analytically) and converges to a strictly positive limit $4 \sin^2(\pi/(2(2k - 1)))$ as $m \rightarrow \infty$, rather than vanishing; the relevant regime for AML

calibration is therefore the dependence on m for moderate cascade depths, not an asymptotic $m \rightarrow \infty$ decay rate. Together with the monotonicity of Theorem 4.4.3 and the stability bound of Theorem 4.4.6, this suggests that a persistent spectral gap at this scale is a natural threshold for the detectability of layering cascades of depth $n - k$.

In the AML feature pipeline of Section 4.8, $\lambda_{\text{gap}}^{(1)}(K_\tau, K_{\tau+1})$ is therefore used as a window-level feature for layering-cascade depth (Table 4.2).

The tridiagonal Schur-complement representation and the resulting spectral characterisation in Theorem 4.5.4 are original to this thesis. While the general spectral properties and stability of persistent Laplacians have been studied, for instance, by Mémoli et al. [2022] and Wei and Wei [2024], we are not aware of a published spectral analysis of the persistent up-Laplacian for path inclusions $P_k \subseteq P_n$ in the form derived below.

Proof sketch. The argument has four steps. First, the up-Laplacian $\Delta_{1,+}^{P_n}$ is reordered into a block form with the P_k -block A , the complementary block D , and a rank-one coupling block C , since only the edges e_{k-1} and e_k share a vertex across the split.

Second, the Schur-complement formula $\Delta_{1,+}^{P_k, P_n} = A - C^\top D^{-1} C$ shows that only the bottom-right entry of A is modified, because C has a single non-zero entry.

Third, D is the invertible edge Gram matrix of the complementary sub-path on $m + 1$ vertices. Its inverse has the explicit closed form $(D^{-1})_{a,b} = \min(a, b)(m + 1 - \max(a, b))/(m + 1)$, giving $(D^{-1})_{1,1} = m/(m + 1)$ and hence a modified corner entry $2 - m/(m + 1) = (m + 2)/(m + 1)$.

Finally, a sine ansatz $v_\ell = \sin(\ell\theta)$ reduces the eigenvalue problem for the resulting tridiagonal matrix to the boundary condition $\sin(k\theta) = \frac{m}{m+1} \sin((k-1)\theta)$. Each of its $k-1$ solutions $\theta \in (0, \pi)$ gives an eigenvalue $\lambda = 2 - 2\cos\theta = 4\sin^2(\theta/2)$, yielding the full spectrum implicitly rather than by a simple closed-form sine grid. $\square$

Theorem 4.5.4 (Spectrum for path inclusions). *Let $P_k \subseteq P_n$ be the inclusion of the k -vertex path as an initial sub-path of P_n , where $2 \leq k \leq n$, and set $m := n - k$.*

If $m = 0$, the eigenvalues of $\Delta_{1,+}^{P_k, P_k}$ are

$$4\sin^2\left(\frac{j\pi}{2k}\right), \quad j = 1, \dots, k-1. \quad (4.33)$$

If $m \geq 1$, $\Delta_{1,+}^{P_k, P_n}$ is the $(k-1) \times (k-1)$ tridiagonal matrix

$$\Delta_{1,+}^{P_k, P_n} = \begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & \frac{m+2}{m+1} \end{pmatrix}, \quad (4.34)$$

whose eigenvalues are

$$\lambda_j = 2 - 2\cos(\theta_j) = 4\sin^2\left(\frac{\theta_j}{2}\right), \quad j = 1, \dots, k-1, \quad (4.35)$$

where $\theta_1, \dots, \theta_{k-1} \in (0, \pi)$ are the $k-1$ solutions of

$$\sin(k\theta) = \frac{m}{m+1} \sin((k-1)\theta). \quad (4.36)$$

Proof. Set $m := n - k$ throughout. We apply the Schur-complement representation (Theorem 4.3.3) to the inclusion $P_k \hookrightarrow P_n$.

Step 1: Block form of $\Delta_{1,+}^{P_n}$. Label vertices $1, \dots, n$ and edges $e_i = \{i, i+1\}$ for $i = 1, \dots, n-1$. The up-Laplacian $\Delta_{1,+}^{P_n} = \partial_1^{P_n} (\partial_1^{P_n})^\top$ is an $(n-1) \times (n-1)$ tridiagonal matrix: its diagonal entries are identically 2 (since every edge contains exactly two vertices), and its off-diagonal entries are -1 whenever two edges share a vertex. Partitioning rows and columns so that the first $k-1$ correspond to $E(P_k) = \{e_1, \dots, e_{k-1}\}$ and the last $n-k$ to $E(P_n) \setminus E(P_k) = \{e_k, \dots, e_{n-1}\}$, the matrix takes the block form

$$\Delta_{1,+}^{P_n} = \begin{pmatrix} A & C^\top \\ C & D \end{pmatrix}, \quad (4.37)$$

where $A \in \mathbb{R}^{(k-1) \times (k-1)}$ is the edge Gram matrix of P_k , $D \in \mathbb{R}^{(n-k) \times (n-k)}$ is the edge Gram matrix of the sub-path $\{k, \dots, n\}$, and $C \in \mathbb{R}^{(n-k) \times (k-1)}$ encodes the coupling at vertex k .

Step 2: The coupling matrix C . The only non-zero entry of C is $(C)_{1,k-1} = -1$, arising from the inner product of edge e_k (which has vertex k as tail, so $(\partial_1^{P_n})_{e_k,k} = -1$) and edge e_{k-1} (which has vertex k as head, so $(\partial_1^{P_n})_{e_{k-1},k} = +1$): the (e_k, e_{k-1}) -entry of $\Delta_{1,+}^{P_n}$ equals $-1 \cdot 1 = -1$. All other entries of C vanish because no other pair of edges from the two blocks shares a vertex. Consequently, $C^\top D^\dagger C$ modifies only the $(k-1, k-1)$ -entry of A .

Step 3: The Schur complement. By Theorem 4.3.3, $\Delta_{1,+}^{P_k, P_n} = A - C^\top D^\dagger C$. The block D is the edge Gram matrix

$$D = \partial_1^{P_{m+1}} (\partial_1^{P_{m+1}})^\top \in \mathbb{R}^{m \times m},$$

of the complementary sub-path on $m+1$ vertices. As an edge Gram matrix of a path, D is positive definite, so $D^\dagger = D^{-1}$. For $1 \leq a, b \leq m$, direct verification shows

$$(D^{-1})_{a,b} = \frac{\min(a,b)(m+1 - \max(a,b))}{m+1},$$

and in particular

$$(D^{-1})_{1,1} = \frac{m}{m+1}. \quad (4.38)$$

Since $C^\top D^{-1}C$ has exactly one non-zero entry, namely $(C^\top D^{-1}C)_{k-1,k-1} = (D^{-1})_{1,1}$, the Schur complement modifies only the bottom-right corner of A :

$$(\Delta_{1,+}^{P_k,P_n})_{k-1,k-1} = 2 - \frac{m}{m+1} = \frac{m+2}{m+1}.$$

All other entries of $\Delta_{1,+}^{P_k,P_n}$ coincide with those of $\Delta_{1,+}^{P_k}$, which proves (4.34).

Step 4: Spectral equation. Let $r := k - 1$. For $\theta \in (0, \pi)$, set $v_\ell := \sin(\ell\theta)$, $\ell = 1, \dots, r$, with $v_0 := 0$. For the interior rows $\ell = 1, \dots, r - 1$, the standard recurrence gives

$$-v_{\ell-1} + 2v_\ell - v_{\ell+1} = (2 - 2\cos\theta)v_\ell.$$

At the final row, the eigenvalue equation reads

$$-v_{r-1} + \frac{m+2}{m+1}v_r = (2 - 2\cos\theta)v_r,$$

which, combined with the interior recurrence, is equivalent to

$$v_{r+1} = \frac{m}{m+1}v_r.$$

Since $r + 1 = k$, this is precisely the boundary condition

$$\sin(k\theta) = \frac{m}{m+1} \sin((k-1)\theta).$$

Conversely, every solution $\theta \in (0, \pi)$ of this equation gives an eigenvector $v = (\sin\theta, \sin 2\theta, \dots, \sin((k-1)\theta))^\top$ of $\Delta_{1,+}^{P_k,P_n}$ with eigenvalue $\lambda = 2 - 2\cos\theta = 4\sin^2(\theta/2)$. As $\Delta_{1,+}^{P_k,P_n}$ is a real symmetric irreducible tridiagonal matrix of size $k - 1$, it has exactly $k - 1$ simple eigenvalues, so (4.36) determines the full spectrum. $\square$

Remark 4.5.5 (Numerical evaluation). For all k and n , the eigenvalues of $\Delta_{1,+}^{P_k,P_n}$ are determined exactly by the $k-1$ roots of the boundary equation (4.36) in $(0, \pi)$. They can thus be computed efficiently from the explicit tridiagonal representation (4.34) without a generic dense Schur-complement computation. The spectrum depends jointly on the retained length k and the extension length $m = n - k$; any AML threshold based on cascade depth should be calibrated from (4.36) rather than from a fixed closed-form grid.

Theorem 4.5.4 provides an explicit tridiagonal representation of the persistent up-Laplacian for path inclusions. Its smallest eigenvalue can be evaluated numerically from the one-dimensional boundary equation (4.36) and used as a window-level feature for layering-cascade depth in the AML pipeline of Section 4.8; the dependence of this gap on the extension length $m = n - k$ can be calibrated for the relevant range of cascade depths.

4.6 Spectral Barcodes and Information-Theoretic Quantities

4.6.1 Classical Barcodes and Their Spectral Extension

Classical persistent barcodes encode, for each homology class, a birth–death interval $[b, d)$ in the filtration parameter and capture purely topological information in the form of persistent Betti numbers and their lifetimes (see, e.g., standard expositions on persistent homology).

By contrast, the full spectrum of the persistent Laplacian $\Delta_q^{K_s, K_t}$ on a simplicial pair (K_s, K_t) contains both this topological information and additional geometric structure: the persistent Hodge theorem implies that persistent Betti numbers are encoded in the harmonic part of the spectrum, while non-zero eigenvalues provide extra multiscale geometrical information beyond barcodes alone [Wei and Wei, 2024, Sec. 3.1].

Definition 4.6.1 (Spectral barcode). Let $\mathcal{K} = \{K_t\}_{t \in T}$ be a filtration in the sense of Definition 3.6.1. The *spectral barcode* of degree q for the pair (K_s, K_t) is the tuple

$$\text{SB}_q(s, t) := (\text{spec}(\Delta_q^{K_s, K_t}), \beta_q^{K_s, K_t}).$$

The multiplicity of the zero eigenvalue of $\Delta_q^{K_s, K_t}$ equals the persistent Betti number $\beta_q^{K_s, K_t}$ by the Betti–nullity correspondence (Theorem 4.2.10). Thus, the zero part of the spectrum records persistent topological information for the pair (K_s, K_t) , whereas the non-zero eigenvalues provide additional spectral information.

Remark 4.6.2 (Spectral barcodes and eigenvectors). While Definition 4.6.1 focuses on eigenvalues, the eigenvectors of the ordinary combinatorial q -Laplacian carry complementary structural information: non-harmonic eigenvectors localise on clusters of q -simplices, while harmonic eigenvectors have support near q -dimensional holes [Wei and Wei, 2024, Sec. 3.3]. In the feature pipeline of Section 4.8 we use only scalar spectral invariants derived from the eigenvalues and eigenvector-based features are left for future work.

4.6.2 Spectral Entropy

The spectral barcode introduced in Definition 4.6.1 records spectral information associated with one filtration pair. However, its dimension grows with the number of simplices and quickly becomes unwieldy if one wishes to use it as a single scalar feature. Motivated by the general observation that eigenvalue distributions of (combinatorial) Laplacians reflect structural and connectivity properties of the underlying complex or graph [Lim, 2020, p. 17][von Luxburg, 2007, pp. 3–6], we now introduce a new scalar invariant: the spectral entropy of a persistent Laplacian. We use this entropy functional here as a problem-specific window-level summary statistic that complements barcode-based and spectral-gap features.

Definition 4.6.3 (Spectral entropy). Let $\text{spec}(\Delta_q^{K,L}) = \{\lambda_1 \leq \dots \leq \lambda_n\}$ be the spectrum of $\Delta_q^{K,L}$, and set $\Lambda := \sum_{k=1}^n \lambda_k$. If $\Lambda > 0$, the *spectral entropy* of the pair (K, L) in degree q is

$$H_{\text{spec}}(K, L, q) := - \sum_{j=1}^n \frac{\lambda_j}{\Lambda} \log \frac{\lambda_j}{\Lambda}, \quad (4.39)$$

where we adopt the convention $0 \log 0 := 0$. Since zero eigenvalues contribute nothing, the effective index set is $I := \{j : \lambda_j > 0\}$ with $|I| = m \leq n$, and the tight upper bound is $\log m$ (Proposition 4.6.5). If $\Lambda = 0$, we set $H_{\text{spec}}(K, L, q) := 0$ by convention, covering both the purely harmonic case ($\beta_q^{K,L} = n$, all eigenvalues zero) and the degenerate case $C_q(K) = \{0\}$.

In other words, the normalised eigenvalues $\{\lambda_j/\Lambda\}_{j \in I}$ are treated as a probability distribution on spectral modes, and H_{spec} measures how evenly spectral energy is spread across these modes. In the AML feature pipeline (Section 4.8), this provides a single scalar indicator of structural heterogeneity within a transaction window and is deliberately constructed to sit alongside the more local geometric information encoded by spectral gaps.

Remark 4.6.4 (Scale invariance of spectral entropy). Set $p_j := \lambda_j/\Lambda$ for $j = 1, \dots, n$ (the same normalised weights used in Definition 4.6.3 and the proof of Proposition 4.6.5). This normalisation makes H_{spec} invariant under a global rescaling of the persistent Laplacian spectrum: replacing every λ_j by $c\lambda_j$ for a constant $c > 0$ leaves each p_j unchanged. Consequently, H_{spec} encodes only the *shape* of the eigenvalue distribution rather than its absolute magnitude.

In the AML pipeline this separation is intentional. The total spectral mass Λ is retained as a separate, scale-sensitive summary of spectral magnitude (Table 4.2), while H_{spec} is used as a unit-free measure of distributional complexity that does not depend on the choice of currency units or normalisation convention.

The normalized weights $p_j := \lambda_j/\Lambda$ form a probability distribution supported on the indices of positive eigenvalues, making H_{spec} an entropy functional on the non-zero spectrum. Since zero eigenvalues, whose multiplicity equals $\beta_q^{K,L}$ by Theorem 4.2.10, do not affect the sum under this convention, the entropy depends solely on the non-zero part of the spectrum.

Proposition 4.6.5 (Bounds on spectral entropy). *Assume $\Lambda > 0$, and let $m \geq 1$ denote the number of strictly positive eigenvalues of $\Delta_q^{K,L}$. Then*

$$0 \leq H_{\text{spec}}(K, L, q) \leq \log m. \quad (4.40)$$

The lower bound is attained if and only if exactly one eigenvalue is strictly positive. The upper bound is attained if and only if all m positive eigenvalues are equal.

Proof. Since $p_j \geq 0$ for all $j \in \{1, \dots, n\}$ and $\sum_{j=1}^n p_j = 1$, the vector $(p_j)_{j \in I}$ forms a probability distribution on the index set I .

Lower bound. For every $j \in I$ we have $p_j \in (0, 1]$, so $\log p_j \leq 0$ and hence $-p_j \log p_j \geq 0$. Summing over I gives $H_{\text{spec}} \geq 0$. Equality holds if and only if $-p_j \log p_j = 0$ for all $j \in I$, which forces $p_j = 1$ for every $j \in I$. Since $\sum_{j \in I} p_j = 1$, this requires $m = 1$.

Upper bound. The elementary inequality $\log t \leq t - 1$ for all $t > 0$ yields, upon setting $t = 1/(mp_j)$ for each $j \in I$,

$$\log \frac{1}{mp_j} \leq \frac{1}{mp_j} - 1.$$

Multiplying by p_j and summing over $j \in I$,

$$\sum_{j \in I} p_j \log \frac{1}{mp_j} \leq \frac{1}{m} \sum_{j \in I} 1 - \sum_{j \in I} p_j = \frac{m}{m} - 1 = 0.$$

Rearranging gives $H_{\text{spec}} = -\sum_{j \in I} p_j \log p_j \leq \log m$. Equality holds if and only if $\log t = t - 1$, that is, $t = 1$, for every $j \in I$, which is equivalent to $p_j = 1/m$ for all $j \in I$, and therefore to all m positive eigenvalues being equal. $\square$

A high value of H_{spec} indicates that spectral energy is evenly spread across all positive eigenvalues at the current filtration step. In the AML application, a high value may be interpreted as evidence that no single positive spectral mode dominates the respective transaction window. Conversely, a low spectral entropy means that spectral energy is concentrated in a small number of dominant modes.

More generally, let $K \hookrightarrow L$ be a pair such that L is unweighted and q -non-branching. If the boundary matrix B_{q+1}^L has full column rank and $B_{q+1}^{L,K} \neq 0$, then the smallest non-zero eigenvalue $\lambda_{\min}^{K,L,q,+}$ of the up-persistent Laplacian is bounded in terms of the connected q -dimensional polyhedra P_j of L :

$$\min_{j \leq h+r} \frac{|\widehat{A}(P_j)|}{|V(P_j)|} \leq \lambda_{\min}^{K,L,q,+} \leq \min_{j \leq h+r} \frac{|A(P_j)|}{|V(P_j)|},$$

where $A(P_j)$ and $\widehat{A}(P_j)$ are as defined in Botnan and Dong [2025, Thm. 4.10]. Thus, under these restrictive structural assumptions, the minimal positive up-persistent eigenvalue admits geometric bounds determined by the constituent polyhedra of the ambient complex L .

No monotonicity property of H_{spec} with respect to filtration extension is assumed in the present work. The entropy is therefore used as a descriptive summary statistic and is interpreted jointly with persistent Betti numbers and spectral-gap features.

In the feature pipeline of Section 4.8, the spectral entropy appears as the window-level feature

$$f_{\tau}^H := H_{\text{spec}}(K_{\tau}, K_{\tau+1}, 1), \quad (4.41)$$

complementing the Betti-number feature f_{τ}^{β} and the spectral-gap features $f_{\tau}^{0,\text{gap}}$ and $f_{\tau}^{1,\text{gap}}$ defined in Section 4.8 below (4.8.1). The four features $(f_{\tau}^{\beta}, f_{\tau}^{0,\text{gap}}, f_{\tau}^{1,\text{gap}}, f_{\tau}^H)$ together form the persistent spectral feature vector $\mathbf{f}_{\tau} \in \mathbb{R}^4$ that is used in the detection algorithm of Section 4.8.

4.7 Filtration Construction from Transaction Networks

Throughout this section, the symmetrised edge weights of the transaction graph are interpreted according to the symmetrisation convention of Section 3.2: the weight w_{ij} of $\{i, j\}$ is one half of the total transaction volume aggregated between i and j in both directions over the observation window. After symmetrisation, transaction direction is deliberately discarded in the simplicial construction. The orientations of simplices are algebraic conventions used to define boundary operators; directed flow information and net-flow imbalances are handled separately by directed Laplacians and degree-based features in Chapter 5.

Throughout this chapter, following Remark 3.1.13, the construction of clique complexes and persistent Laplacians is applied to induced subgraphs $D[U]$ of the full transaction graph D , where U denotes a set of accounts pre-selected by upstream detection mechanisms (e.g., rule-based alerts, classical graph features, or GNN models). In this architecture, the persistent Laplacian acts as a high-precision operator on candidate fraud subnetworks, rather than on the entire payment network, which mitigates the combinatorial explosion of cliques in dense hubs.

4.7.1 Data Basis: The AMLworld Transaction Graph

Our empirical evaluation is based on the AMLworld synthetic transaction dataset [Altman et al., 2024], which models a realistic multi-bank transaction network using a parameterised virtual-world simulator and provides complete simulated ground-truth money-laundering labels. Each record specifies a sender account, a receiver account, a transaction amount, a timestamp, and a binary fraud label.

Definition 4.7.1 (AMLworld transaction graph). Let $\mathcal{T} = \{(u_i, v_i, w_i, t_i, \ell_i)\}_{i=1}^N$ be the AMLworld transaction list, where $u_i, v_i \in V$ are account identifiers, $w_i > 0$ is the transaction amount, $t_i \in \mathbb{R}$ is the timestamp, and $\ell_i \in \{0, 1\}$ is the fraud label.

The *AMLworld transaction digraph* is the directed weighted pseudograph $G = (V, E, w)$ with vertex set

$$V = \{u_i\}_{i=1}^N \cup \{v_i\}_{i=1}^N.$$

Its arc multiset is

$$E := \{a_i\}_{i=1}^N, \quad a_i := (u_i, v_i),$$

and its weight function is defined on arcs by

$$w : E \rightarrow \mathbb{R}_{>0}, \quad w(a_i) := w_i.$$

Because G is directed and may contain multi-arcs, it falls within the directed multigraph framework of Definition 3.1.9 (Section 3.2). In the empirical pipeline of Chapter 5, this abstract construction is instantiated on the HI-Small subset of the AMLWORLD dataset as the directed weighted multigraph

$$D = (V_{\text{sub_final}}, A_{\text{sub}}, w_{\text{wins}}),$$

where $V_{\text{sub_final}}$ and A_{sub} are obtained via the one-hop and two-hop induced subgraph construction of Section 5.2, and w_{wins} denotes the EUR-converted, winsorised transaction volumes introduced in Section 5.2.7. The laundering indicator **Is_Laundering** and the node-level fraud roles (**Fraud_sender**, **Fraud_receiver**, **Fraud_intermediate**) are stored as metadata on $V_{\text{sub_final}}$ and A_{sub} (Section 5.2.8) and do not enter the graph or topological constructions.

Symmetrisation via G_{sym} (Definition 3.2.13) converts G into an undirected weighted graph on which simplicial complexes and Laplacians are well-defined; loops are excluded from G_{sym} in accordance with Remark 3.1.2.

For distinct accounts $u, v \in V$, define the directionally aggregated transaction volume by

$$\bar{w}(u, v) := \sum_{i: u_i=u, v_i=v} w_i.$$

The symmetrised edge weight is then

$$w_{\text{sym}}(\{u, v\}) := \frac{1}{2}(\bar{w}(u, v) + \bar{w}(v, u)),$$

representing the symmetrised average of the total directional transaction volumes, in accordance with the convention of Section 3.2. This produces the undirected weighted graph used for the clique-complex construction and persistent-Laplacian analysis.

4.7.2 Temporal Filtration via Sliding Windows

Definition 4.7.2 (Sliding-window graph). Fix a window length $W > 0$ and a step size $\delta > 0$. For each window index $\tau \in \mathbb{N}_0$, the *sliding-window directed graph* is

$$G^\tau := \left(V, ((u_i, v_i, w_i))_{\tau\delta \leq t_i < \tau\delta + W}, w \right). \quad (4.42)$$

Its symmetrisation G_{sym}^τ and the corresponding clique complex $K_\tau := \text{Cl}(G_{\text{sym}}^\tau)$ (Definition 3.4.4) define the *temporal snapshot* at window τ . If G_{sym}^τ contains no edges, K_τ is taken to be the empty complex and the corresponding persistent Laplacian is the zero operator.

Rather than accumulating complexes indefinitely via the global sequence $K_0 \subseteq K_0 \cup K_1 \subseteq \dots$, our architecture isolates step-to-step structural dynamics. For each consecutive pair of sliding windows, the inclusion $K_\tau \hookrightarrow K_\tau \cup K_{\tau+1}$ defines a valid simplicial pair in the sense of Definition 3.6.6. We then compute the up-persistent Laplacian $\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$ and record its spectrum, yielding a time series of localized spectral snapshots across the observation period. The persistent Betti number $\beta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$ (Theorem 4.2.10) measures the number of one-dimensional homology classes present in K_τ that persist in $K_\tau \cup K_{\tau+1}$. It provides the topological feature $f_\beta(\tau)$ defined in (4.44).

The hyperparameters W and δ control the temporal resolution of the filtration: a large W smooths over short-lived structures but may conflate fraud rings active in different periods; a small δ increases the number of spectral snapshots but raises the computational cost due to the matrix diagonalisation required for the persistent Laplacian spectrum (Algorithm 2, Section 4.3.4).

4.7.3 Volume-Based Filtration

Definition 4.7.3 (Volume filtration). For a strictly decreasing threshold sequence $\theta_0 > \theta_1 > \dots > \theta_T \geq 0$, the *volume filtration* is the nested family $\{K_{\theta_s}\}_{s=0}^T$, where

$$K_{\theta_s} := \text{Cl}(G_{\text{sym}}[\{e \in E(G_{\text{sym}}) : w_{\text{sym}}(e) \geq \theta_s\}]) \quad (4.43)$$

is the clique complex of the edge-induced spanning subgraph of G_{sym} retaining only edges with symmetrised weight at least θ_s . Since $\theta_s > \theta_{s+1}$ implies $K_{\theta_s} \subseteq K_{\theta_{s+1}}$ for all s , the family $\{K_{\theta_s}\}_{s=0}^T$ is a simplicial filtration in the sense of Definition 3.6.1.

Here $w_{\text{sym}}(e)$ denotes the symmetrised average transaction volume on the edge $e = \{i, j\}$, i.e., one half of the total transaction amount aggregated between i and j in both directions over the observation window, as specified by the symmetrisation convention in Section 3.2. For the volume-based filtration used in this chapter, we adopt the reparametrisation introduced in Section 3.6.2: a decreasing sequence of volume thresholds (θ_s) induces an increasing sequence of filtration parameters (t_s) with $K_{t_s} = K_{\theta_s}$ and $t_s < t_\ell \Rightarrow K_{t_s} \subseteq K_{t_\ell}$. This ensures that all persistent homology and persistent Laplacian constructions apply with their standard inclusion assumptions.

Remark 4.7.4. The non-strict inequality $w_{\text{sym}}(e) \geq \theta_s$ in (4.43) is the defining convention of this thesis and must be applied consistently wherever transaction weights are compared to a threshold. Both conventions preserve the filtration property $K_{\theta_s} \subseteq K_{\theta_{s+1}}$ when $\theta_s > \theta_{s+1}$. However, they assign different birth indices to edges whose weight equals a threshold value exactly: with $\geq$, an edge with $w_{\text{sym}}(e) = \theta_s$ is born at step s ; with $>$, it would be born at step $s+1$. We adopt $\geq$ throughout so that K_{θ_s} is the *largest* subcomplex consistent with threshold θ_s , and the map $\theta \mapsto K_\theta$ is right-continuous. Decreasing θ admits lower-volume edges and can reveal scatter-gather and gather-scatter structures, which are typical smurfing topologies [Deprez et al., 2025a]. An isolated fan-in/fan-out structure is a tree and therefore has $\beta_1 = 0$. Such structures contribute non-trivial one-dimensional homology only if additional cross-links between branches create cycles. When a boundary-closing edge e satisfies

$$\theta_s > w_{\text{sym}}(e) \geq \theta_{s+1},$$

it is absent from K_{θ_s} and present in $K_{\theta_{s+1}}$; this may create a new cycle. The persistent Betti number $\beta_1^{K_{\theta_s}, K_{\theta_{s+1}}}$ counts the one-dimensional homology classes of K_{θ_s} that persist in $K_{\theta_{s+1}}$ (Theorem 4.2.10).

The two filtration strategies are complementary and may be combined: the temporal filtration tracks *when* cyclic structures emerge, while the volume filtration tracks *at what transaction size* they become topologically significant.

The temporal construction is applied to the window-specific symmetrised graphs G_{sym}^τ , whereas the volume filtration is applied to the symmetrised graph G_{sym} of the selected observation period. Their respective features can be combined in the detection pipeline of Section 4.8 when both constructions are implemented for the same candidate subnetwork.

The persistent spectral gap $\lambda_{\text{gap}}^{(1)}(K_{\theta_s}, K_{\theta_{s+1}})$ provides a volume-scale higher-dimensional spectral-gap feature. Unlike the Fiedler value, which is associated with the 0-Laplacian, it is interpreted here as a descriptor of the non-zero persistent 1-spectrum.

4.8 Spectral Features for AML Detection

Throughout this section, we write $H_{\text{spec}}(\tau) := H_{\text{spec}}(K_\tau, K_\tau \cup K_{\tau+1}, 1)$ for the degree-1 spectral entropy at filtration step τ , consistently with Definition 4.6.3.

4.8.1 Window-Level Feature Definitions

For each window pair $(K_\tau, K_\tau \cup K_{\tau+1})$, we extract the following scalar features from the spectrum of $\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$. Throughout, eigenvalues are ordered non-decreasingly and counted with multiplicity; by the Betti–nullity correspondence (Theorem 4.2.10), the first $\beta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$ eigenvalues of $\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$ are zero, so provided that a positive eigenvalue exists, $\lambda_{\beta_1^{K_\tau, K_\tau \cup K_{\tau+1}}}(\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}})$ denotes the smallest positive eigenvalue.

Definition 4.8.1 (Spectral AML features).

$$f_\beta(\tau) := \beta_1^{K_\tau, K_\tau \cup K_{\tau+1}} = \text{nullity}(\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}), \quad (4.44)$$

$$\mathcal{F}_\beta(\tau) := f_\beta(\tau) - f_\beta(\tau - 1), \quad \tau \geq 1, \quad (4.45)$$

$$f_{\text{gap}}^{(0)}(\tau) := \lambda_{\beta_0^{K_\tau, K_\tau \cup K_{\tau+1}}}(\Delta_0^{K_\tau, K_\tau \cup K_{\tau+1}}), \quad (4.46)$$

$$f_{\text{gap}}^{(1)}(\tau) := \lambda_{\beta_1^{K_\tau, K_\tau \cup K_{\tau+1}}}(\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}), \quad (4.47)$$

$$f_H(\tau) := H_{\text{spec}}(K_\tau, K_\tau \cup K_{\tau+1}, 1), \quad (4.48)$$

$$f_\Lambda(\tau) := \Lambda_1(K_\tau, K_\tau \cup K_{\tau+1}) := \sum_k \lambda_k(\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}). \quad (4.49)$$

If $\Delta_q^{K_\tau, K_\tau \cup K_{\tau+1}}$ has no positive eigenvalue, we set $f_{\text{gap}}^{(q)}(\tau) := 0$.

The same definitions apply verbatim with τ replaced by a volume filtration index s and $K_\tau \cup K_{\tau+1}$ replaced by K_t , yielding the feature vector $\mathbf{f}(s)$ used in Section 4.8.3 and Section 4.8.4.

Remark 4.8.2. Eigenvalues of $\Delta_0^{K,L}$ are indexed from 0. For a classical graph Laplacian of a connected graph, $\lambda_0 = 0$ and λ_1 is the Fiedler value. More generally, the multiplicity of the eigenvalue 0 equals the number of connected components; thus, the smallest positive eigenvalue is indexed by the nullity of the Laplacian. In the classical graph-Laplacian setting, the Fiedler value is positive if and only if the graph is connected [Spielman, 2025, Sec. 1.5.4, pp. 16–17].

Remark 4.8.3. The notation $\mathcal{F}_\beta$ distinguishes this scalar *feature score* from the persistent Laplacian operator $\Delta_q^{K,L}$ (Definition 4.2.4). Specifically, $\mathcal{F}_\beta(\tau) \in \mathbb{Z}$ is computed as

$$\mathcal{F}_\beta(\tau) := \text{nullity}(\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}) - \text{nullity}(\Delta_1^{K_{\tau-1}, K_{\tau-1} \cup K_\tau}),$$

i.e. the change in the nullity of the persistent 1-Laplacian over one filtration step of width δ .

Temporal burst features $\mathcal{F}_\beta$, defined via the sliding-window snapshot (Section 4.7.2), provide an additional conceptual link to smurfing cascades (Section 2.3.5), but are documented in Chapter 5 as an empirical limitation of the present implementation and are therefore not included in the account-level vector $\mathbf{x}(a)$.

The four features (4.44), (4.46), (4.47), and (4.48) constitute the *persistent spectral feature vector*

$$\mathbf{f}(\tau) := (f_\beta(\tau), f_{\text{gap}}^{(0)}(\tau), f_{\text{gap}}^{(1)}(\tau), f_H(\tau))^\top \in \mathbb{R}^4.$$

Their theoretical properties follow directly from the results of Sections 4.2–4.6: f_β is an integer-valued topological invariant (Theorem 4.2.10); $f_{\text{gap}}^{(0)}$ and $f_{\text{gap}}^{(1)}$ are non-negative spectral features. The monotonicity theorem applies to the eigenvalues of the up-persistent Laplacian under nested filtration extensions. The feature f_H is a scalar summary of the positive persistent spectrum, as defined in Definition 4.6.3.

4.8.2 Mapping to AML Patterns

Table 4.2 closes the circle between the criminological typologies of Chapter 2 and the spectral features defined above. For each feature, the table records the theoretical result that justifies its use as an indicator, the AML pattern it targets, and the relevant section of Chapter 2.

4.8.3 Account-Level Aggregation

Volume-level feature vectors are aggregated to the account level by summing contributions from all filtration levels in which account a participates. Here $V(K_s)$ denotes the vertex set of K_s , i.e. the set of accounts that appear in at least one transaction retained at filtration level s .

Definition 4.8.4 (Account feature vector). For account $a \in V$, define

$$\mathbf{x}(a) := \sum_{s: a \in V(K_s)} \mathbf{f}(s) \in \mathbb{R}^4. \quad (4.50)$$

The sum in (4.50) accumulates spectral evidence across all filtration levels in which account a is active. Accounts that participate repeatedly in filtration levels with non-zero persistent spectral features accumulate the corresponding feature contributions in $\mathbf{x}(a)$. The empirical differences between fraud-labelled and background accounts are assessed descriptively in Chapter 5. The vector $\mathbf{x}(a)$ serves as an account-level spectral representation for the descriptive evaluation in Chapter 5.

4.8.4 Algorithmic Pipeline for Chapter 5

The full computational pipeline, based on Algorithm 1 (Section 4.3.3) and Algorithm 2 (Section 4.3.4), proceeds as follows.

Table 4.2: Spectral AML features: theoretical grounding and pattern correspondence. Section references are to Chapter 2 and theorem references are to this chapter.

Feature	Symbol	Spectral interpretation	AML pattern	Theorem
Pers. Betti-1	f_β	Number of independent 1-cycles persisting from K_τ to $K_\tau \cup K_{\tau+1}$. 0 if no 1-cycle persists across this step	Cyclic structures (Sec. 2.3.1)	Thm. 4.2.10
Betti increment	$\mathcal{F}_\beta$	Positive and negative values indicate changes in the number of one-dimensional homology classes persisting across consecutive window pairs	Temporal bursts (Sec. 2.3.5)	Thm. 4.2.10
Spectral gap (0)	$f_{\text{gap}}^{(0)}$	Fiedler value of $\Delta_0^{K_\tau, K_\tau \cup K_{\tau+1}}$, small values certify bottlenecks separating clusters	Layering cascades (Sec. 2.3.3)	Thm. 4.4.9
Spectral gap (1)	$f_{\text{gap}}^{(1)}$	Smallest positive eigenvalue of $\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$; quantifies geometric robustness of cycles. For non-branching paths, given implicitly by the boundary equation (4.36) (Theorem 4.5.4); decreases monotonically in cascade depth $m = n - k$ for fixed k , but converges to a strictly positive limit rather than vanishing as $m \rightarrow \infty$	Cycle persistence, fraud-ring depth (Sec. 2.3.1)	Thm. 4.5.4
Spectral entropy	f_H	Entropy of positive eigenvalues; high values indicate heterogeneous topology, low values dominance of a single mode	Mixed structuring (Sec. 2.3.2)	Prop. 4.6.5, Def. 4.6.3
Spectral mass	f_Λ	Total spectral weight $\Lambda_q = \sum_k \lambda_k(\Delta_q^{K_\tau, K_\tau \cup K_{\tau+1}})$; provides a scale-sensitive summary of spectral magnitude complementary to f_H	Placement / absolute volume (Sec. 2.2.1)	Def. 4.6.3

1. **Data ingestion.** Read the AMLworld dataset [Altman et al., 2024]; extract the transaction list $\mathcal{T}$ and the ground-truth fraud labels on edges and accounts (Section 5.2).
2. **Working graph and volume filtration.** Construct the directed weighted multigraph $D = (V, A, w)$ and remove self-loops. From D we either aggregate parallel arcs to obtain a weighted digraph (Section 5.3), or symmetrise the transaction network to obtain the undirected graph G_{sym} (Section 5.4), depending on whether we follow the directed-graph or the simplicial/spectral analysis path. On the pattern-centred working subgraph G_{work} (Definition 5.4.1), derive a strictly decreasing sequence of volume thresholds $\{\theta_s\}_{s=0}^S$ from edge-weight quantiles and construct the volume-based filtration $(K_s)_{s=0}^S$ with $K_s := \text{Cl}(G_{\text{sym}}[\theta_s])$ in Section 5.4.3.
3. **Simplicial complex.** For each filtration level s , build the clique complex $K_s := \text{Cl}(G_{\text{sym}}[\theta_s])$ (Definition 3.4.4) and compute its chain groups and boundary operators $\partial_0^{K_s}, \partial_1^{K_s}$ following Section 3.4.
4. **Boundary matrices and Schur blocks.** For each filtration pair (K_s, K_t) with $s < t$ used in Chapter 5, assemble the boundary matrix $\partial_1^{K_t}$ of the larger complex, index its rows and columns by the simplices of K_s and K_t , and extract the blocks B_{11}, B_{12}, B_{22} according to Algorithm 1, Steps 1–2 (Section 4.3.3).
5. **Persistent Laplacian and spectrum.** Apply Algorithm 2 (Section 4.3.4) to each pair (K_s, K_t) to construct the persistent 1-Laplacian $\Delta_1^{K_s, K_t}$ via the Schur-complement representation, and compute its full spectrum on the working graph (Section 5.5).
6. **Feature extraction (window level).** For each filtration pair (K_s, K_t) , compute the persistent spectral feature vector $\mathbf{f}(s) \in \mathbb{R}^4$ as defined in (4.44), (4.46), (4.47), and (4.48): the persistent Betti count f_β , the 0- and 1-spectral gaps $f_{\text{gap}}^{(0)}, f_{\text{gap}}^{(1)}$, and the spectral entropy f_H (Definition 4.8.1).
7. **Account-level aggregation.** Map the volume-level feature vectors $\mathbf{f}(s)$ back to individual transactions in $\mathcal{T}$ via their birth step, and aggregate contributions over all filtration levels incident to each account a according to (4.50): for each a , $\mathbf{x}(a)$ is obtained by summing and normalising the feature contributions of all edges for which a appears as sender or receiver, yielding an account-level spectral feature vector.
8. **Scoring and descriptive evaluation.** Combine $\mathbf{x}(a)$ with classical graph indicators such as fan-in/fan-out (Section 2.3.2) and volume-based filtration statistics (Section 4.7.3) to obtain interpretable AML anomaly scores at the account level. Chapter 5 evaluates these scores descriptively by comparing their distributions across fraud-labelled and background accounts and reporting summary statistics and rank-based measures (e.g. Spearman correlations), rather than training a separate classification model.

The betti-nullity correspondence, the up-persistent eigenvalue monotonicity result, and the persistent Cheeger bound provide theoretical motivation for components of the feature vector $\mathbf{x}(a)$. The empirical behaviour of the aggregated representation is evaluated descriptively in Chapter 5. In particular, the persistent spectral feature vector $\mathbf{x}(a) \in \mathbb{R}^4$ translates the criminological typology of Chapter 2 into a mathematically grounded numerical signature. Each component targets a distinct structural phenomenon: f_β counts the independent cyclic flow patterns that are the hallmark of placement rings (Section 2.3.1); $f_{\text{gap}}^{(0)}$ certifies the bottleneck separations exploited in layering cascades via the persistent Cheeger-type interpretation (Section 2.3.3, Theorem 4.4.9), $f_{\text{gap}}^{(1)}$ records the smallest positive eigenvalue of the persistent 1-Laplacian and is used as a spectral descriptor of persistent one-dimensional structure (Section 4.5) and f_H captures the spectral heterogeneity of mixed or structuring schemes in which several cyclic patterns co-exist at comparable spectral energy (Section 2.3.2, Section 4.6.2). Temporal burst features $\mathcal{F}_\beta$, defined via the sliding-window snapshot (Section 4.7.2), provide an additional conceptual link to smurfing cascades (Section 2.3.5), but are documented in Chapter 5 as an empirical limitation of the present implementation and are therefore not included in the account-level vector $\mathbf{x}(a)$.

Together, these spectral-topological features provide what classical graph indicators cannot: a topological and spectral characterisation of the transaction network that is interpreted jointly with the volume-filtration scales, interpretable in regulatory terms, and tractable on the pattern-centred working subgraph used in Chapter 5. The empirical evaluation of this claim (including Nullity checks, path-spectrum calibration, and account-level aggregation effects) is the subject of Chapter 5.

Chapter 5

Results and Discussion

Terminology. Throughout this chapter we use AML pattern terminology such as cycles and rings, fan-in and fan-out structures, bipartite patterns, layering cascades, and strongly connected components. Formal definitions of these terms are given in Chapter 2 (Sections 2.2–2.3) and Chapter 3 (Section 3.3.2). We briefly recall that fan-in structures correspond to vertices of high in-degree, fan-out structures to vertices of high out-degree, and bipartite patterns denote two-partite sender–receiver blocks in the AMLworld generator.

5.1 Structure of this chapter

This chapter applies the theoretical framework developed in Chapter 4 to the AMLWORLD transaction data set. Our goal is to validate the spectral and topological concepts on real-world anti-money-laundering data and to use them to characterise fraud-related structures and account types.

We proceed in several steps. Section 5.2 describes the data and the construction of the transaction graphs. In Section 5.3 we analyse the directed network in terms of strongly connected components and directed degree patterns. Sections 5.4 and 5.5 introduce the simplicial framework, filtrations, and persistent Laplacians, and extract window-level spectral features. Section 5.6 aggregates these features to the account level and discusses emergent typologies. Finally, Section 5.7 summarises our findings in terms of the research questions stated in Chapter 1.

Two complementary graph perspectives In our empirical analysis we work with two closely related, but conceptually different representations of the AMLWORLD transaction network.

Approach A: directed

We first construct the aggregated, weighted digraph D_{simple} (Section 5.3). This directed network preserves edge orientations and is used for purely combinatorial structure diagnostics such as strongly connected components, fan-in/fan-out distributions, and directed degree patterns.

Approach B: symmetrised

In parallel, we consider the symmetrised version G_{sym} (and its working variant G_{work}) introduced in Section 5.4. This undirected graph is the basis for our clique complex, homology computations, persistent Laplacians, and the associated spectral features.

The two perspectives are consistent in the sense that dense, recurrent transaction regions that appear as strongly connected components in D_{simple} typically correspond to cyclic or ring-like regions in the simplicial/spectral picture of G_{sym} . However, a full comparative study of the two approaches lies beyond the scope of this thesis. Here we deliberately use the directed view for combinatorial structure diagnostics and the symmetrised view for topological and spectral analyses.

5.2 Data preparation and graph construction

The empirical analysis is based on the synthetic *HI-Small* subdataset of the AMLWORLD dataset by [Altman et al. \[2024\]](#). The raw data are modelled as a directed weighted multigraph $D(V, A, w)$ in the sense of Definition 3.1.1, where vertices represent bank accounts and arcs represent individual transactions. Data preparation proceeds in sequential steps, which are detailed below.

5.2.1 Identification of the illicit vertex set V_{ill}

From the patterns file `HI-Small_Patterns.txt` all account numbers appearing in at least one BEGIN/END laundering block are extracted. This set $V_{\text{ill}} \subseteq V$ contains all vertices on which the topologically relevant structures (cycles, fan-in/fan-out hubs, strongly connected components) are defined and which therefore must not be removed by any subset construction. In total, 370 pattern blocks over eight distinct typologies are identified, as summarized in Table 5.1.

Table 5.1: Number of laundering pattern blocks by type in the HI-Small dataset.

Pattern type	Blocks
BIPARTITE	49
CYCLE	54
FAN-IN	40
FAN-OUT	48
GATHER-SCATTER	51
RANDOM	41
SCATTER-GATHER	44
STACK	43
Total	370

The resulting fraud-vertex set contains $|V_{\text{ill}}| = 3,170$ distinct account identifiers.

5.2.2 Fraud-role annotation

For each vertex $v \in V_{\text{ill}}$, three binary operational roles within the laundering blocks are determined:

- **Fraud_sender**: Vertex appears as a sender in at least one block (2,782 vertices).
- **Fraud_receiver**: Vertex appears as a receiver in at least one block (2,882 vertices).
- **Fraud_intermediate**: Vertex appears in at least one block both as sender and receiver, acting as a man-in-the-middle (2,494 vertices).

The aggregate flag **Fraud** = 1 holds for all 3,170 vertices in V_{ill} .

5.2.3 Loading transactions and removal of self-loops

The raw transactions from `HI-Small_Trans.csv` comprise 5,078,345 entries. By Definition 3.1.1, self-loops ($u \rightarrow u$) are excluded from the multigraph. The 591,212 corresponding loop entries are removed, leaving 4,487,133 valid transactions for the network expansion.

5.2.4 One-hop filter and induced subgraph

Since the full vertex set $|V| \approx 3.9 \cdot 10^6$ is computationally intractable for the direct evaluation of persistent Laplacians, an induced subgraph $D[V_{\text{sub}}]$ is constructed. As a first step, the one-hop neighborhood extension

$$V_{\text{sub}}^{(1)} = \bigcup_{v \in V_{\text{ill}}} N^1(v) \supseteq V_{\text{ill}}$$

is formed by retaining all arcs $(u \rightarrow v) \in A$ such that $u \in V_{\text{ill}}$ or $v \in V_{\text{ill}}$. This yields 100,567 transactions and $|V_{\text{sub}}^{(1)}| = 12,842$ vertices, of which 3,170 are fraud vertices and 9,672 are benign one-hop neighbours. A formal consistency check confirms the complete containment $V_{\text{ill}} \subseteq V_{\text{sub}}^{(1)}$.

5.2.5 Two-hop extension to target budget

Since the one-hop size $|V_{\text{sub}}^{(1)}| = 12,842$ falls short of the target computational budget of 30,000 nodes, the vertex set is expanded via a two-hop extension. From the 75,606 available two-hop candidate vertices, exactly 17,158 nodes are drawn uniformly at random with a fixed pseudorandom seed (42) and appended to the vertex set:

$$V_{\text{sub_final}} = V_{\text{sub}}^{(1)} \cup \tilde{V}^{(2)}, \quad |\tilde{V}^{(2)}| = 17,158, \quad |V_{\text{sub_final}}| = 30,000.$$

By construction, the topological guarantee $V_{\text{ill}} \subseteq V_{\text{sub_final}}$ remains strictly preserved.

5.2.6 Final induced subgraph topology

The final arc set A_{sub} is established by restricting the network to transactions whose endpoints both reside within the bounded vertex set:

$$A_{\text{sub}} = \{(u, v) \in A \mid u, v \in V_{\text{sub_final}}, u \neq v\}.$$

This restriction yields $|A_{\text{sub}}| = 299,268$ total transactions, of which 3,399 are explicitly flagged as laundering events. This corresponds to an illicit edge rate of 1.14%, which represents a significant concentration compared to the 0.12% base rate in the raw dataset. Table 5.2 documents the distribution of cross-sectional arc types.

Table 5.2: Arc types in the induced subgraph $D[V_{\text{sub_final}}]$.

Arc type	Count	Share
neighbor_to_neighbor	198,701	66.40 %
fraud_to_neighbor	56,556	18.90 %
neighbor_to_fraud	38,876	12.99 %
fraud_to_fraud	5,135	1.72 %
Total	299,268	100.00 %

Arc types are defined purely by the endpoint roles (Fraud vs. neighbor) and do not coincide with the laundering label `Is_Laundering=1` on the transaction itself. In particular, a `fraud_to_fraud` arc denotes a transfer between two vertices involved in at least one laundering block, but the arc need not be explicitly flagged as a laundering event.

5.2.7 Feature engineering and currency conversion

To standardize the edge weights, all transaction amounts are converted to Euro (EUR). Exchange rates are parameterized as mid-market values as of 5 July 2026, extracted from the public FX validation tables of the German Banking Association (Währungsrechner, Bankenverband).¹ The functional mapping $r_{\text{FX}}(\cdot)$ from the raw currency code to the EUR multiplier is specified as:

$$\begin{aligned} r_{\text{FX}}(\text{EUR}) &= 1.0000, & r_{\text{FX}}(\text{USD}) &= 0.8700, & r_{\text{FX}}(\text{GBP}) &= 1.1700, & r_{\text{FX}}(\text{CNY}) &= 0.1287, \\ r_{\text{FX}}(\text{JPY}) &= 0.0054, & r_{\text{FX}}(\text{INR}) &= 0.0092, & r_{\text{FX}}(\text{RUB}) &= 0.0114, & r_{\text{FX}}(\text{CAD}) &= 0.6154, \\ r_{\text{FX}}(\text{AUD}) &= 0.6056, & r_{\text{FX}}(\text{MXN}) &= 0.0500, & r_{\text{FX}}(\text{BRL}) &= 0.1690, & r_{\text{FX}}(\text{CHF}) &= 1.0910, \\ r_{\text{FX}}(\text{ILS}) &= 0.2914, & r_{\text{FX}}(\text{SAR}) &= 0.2327. \end{aligned}$$

¹<https://bankenverband.de/services/waehrungsrechner>, accessed 5 July 2026.

Bitcoin transactions are treated via a synthetic scaling unit with $r_{\text{FX}}(\text{BTC-Unit}) = 1.0000$, mapping one nominal BTC unit directly to one effective Euro. This parameter choice is consistent with the underlying data generator design and prevents extreme weight imbalances that would otherwise distort the filtration scales if real-world exchange rates ($\approx 5.5 \times 10^4 \text{ EUR}^2$) were applied.

The EUR-denominated transaction weight $w_{\text{EUR}}(u, v)$ for an arc (u, v) with raw amount $a(u, v)$ is given by:

$$w_{\text{EUR}}(u, v) = a(u, v) \cdot r_{\text{FX}}(\text{cur}(u, v)).$$

To regularize the multi-scale filtration thresholds and mitigate the impact of massive outliers on the quantile boundaries (cf. Definition 4.7.3), the edge weights are winsorized at the 99th percentile:

$$w_{\text{wins}}(u, v) = \min(w_{\text{EUR}}(u, v), p_{0.99}), \quad p_{0.99} = 2,806,368.02 \text{ EUR}.$$

This monotonic transformation retains the structural ordering of the network layers while ensuring that macroscopic outliers do not compress the filtration resolution. For alternative robustness assessments, a log-scaled weight configuration $w_{\log}(u, v) = \log(1 + w_{\text{wins}}(u, v))$ is maintained. Unless stated otherwise, w_{wins} serves as the canonical weight function for all downstream topological constructions, persistent homology metrics, and persistent Laplacians $\Delta_q^{K,L}$.

5.2.8 Fraud labels on nodes versus edges

A key methodological premise of this study is the conceptual distinction between nodes representing crime roles and edges representing events.

On the node level, the vertex set V_{ill} and the binary flags in `account_info.csv` represent a long-term structural classification. Once an account is identified as an active sender, receiver, or intermediate within a parsed laundering block from `HI-Small_Patterns.txt`, it is flagged as a fraud node (`Fraud` = 1) for the entirety of the topological analysis, independent of whether its other incident arcs are legitimate or illicit. The final vertex set $V_{\text{sub_final}}$ contains exactly 30,000 accounts, comprising 3,170 fraud nodes (10.57%) and 26,830 benign neighbor nodes.

On the edge level, the field `Is_Laundering` in `transactions_30kvertices.csv` represents a discrete transactional label. In the induced graph, there are 3,399 transactions marked with `Is_Laundering` = 1 (1.14% of A_{sub}). These edge labels are heavily concentrated within the structural core: 3,199 of these laundering transactions connect two fraud nodes (`fraud_to_fraud`), representing 94.12% of all laundering edges. The remaining fraud edges are distributed across the mixed and neighbor categories: 24 are `fraud_to_neighbor`, 17 are `neighbor_to_fraud`, and 159 are `neighbor_to_neighbor`.

Crucially, while 62.3% of the total `fraud_to_fraud` edges (5,135 in total) are actively executing laundering patterns, a significant portion (37.7%) represents non-laundering, standard financial traffic between the same crime actors. Conversely, `neighbor_to_neighbor` laundering arcs indicate transactions flagged by the generator where the underlying endpoints fell outside the explicit pattern blocks. To verify the consistency of the network pipeline, we check that every endpoint classified as “fraud”

²Cf. BTC/EUR market quotes on 5 July 2026.

or “neighbor” in the edge-level `Transaction_Type` matrix has a matching node-level flag in V_{ill} and $V_{\text{sub_final}}$; no edge-level criminological role is assigned to a vertex that is not consistently labelled at the node level.

5.3 Construction of the transaction graph

5.3.1 Directed weighted multigraph D

Based on the arc set A_{sub} constructed in Section 5.2, the transaction graph

$$D = (V_{\text{sub_final}}, A_{\text{sub}}, w)$$

is realised as a finite directed weighted multigraph in the sense of Definition 3.1.1. Each transaction is modelled as an individual arc; parallel arcs $u \rightarrow v$ (multiple transactions between the same pair of accounts) are retained, since aggregating them would distort the thresholds $\{K_t\}_{t \in T}$ for the filtrations, see Chapter 4.

Remark 5.3.1 (Edge Preservation). In this structural and descriptive stage, parallel edges are explicitly preserved to maintain the microstructural integrity of the transaction network. Each edge represents an individual financial transaction with unique temporal and volumetric attributes. Aggregating these edges at this point would distort key network metrics—such as transaction frequency and temporal gaps between successive transfers—which are essential for a comprehensive empirical description of the financial flows.

The edge weight is the winsorised transaction volume $w(u, v) = w_{\text{wins}}(u, v) \in \mathbb{R}_{\geq 0}$ (see Section 5.2.7); timestamps are kept as attributes, as they underlie the temporal filtration in Chapter 4. The fraud label `Is_Laundering` is stored solely as metadata and does not enter the graph or topological constructions.

5.3.2 Descriptive graph statistics

For descriptive statistics and structural analysis (such as degree distributions and connected components), we pass from the transaction multigraph $D = (V_{\text{sub_final}}, A_{\text{sub}}, w)$ to a collapsed digraph D_{simple} by identifying all parallel arcs between the same ordered vertex pair. In line with the discussion in Chapter 3, multiple transactions between the same pair of accounts are thus collapsed into a single unweighted edge, yielding a simplified structural view of the connectivity pattern. D_{simple} has the same vertex set as D and contains an arc (u, v) if and only if there exists at least one transaction from u to v in A_{sub} :

$$V(D_{\text{simple}}) = V_{\text{sub_final}}, \quad A(D_{\text{simple}}) = \{(u, v) \in V^2 : \exists a \in A_{\text{sub}} \text{ with } \text{tail}(a) = u, \text{ head}(a) = v\}.$$

Edge weights and multiplicities are ignored in D_{simple} , which is used exclusively for purely combinatorial quantities such as densities and node degrees. All multi-scale weighted constructions (filtrations,

persistent homology, and persistent Laplacians) are carried out strictly on the original multigraph D with the winsorised weights $w(u, v) = w_{\text{wins}}(u, v)$.

Table 5.3 summarises the key structural quantities.

Table 5.3: Structural characteristics of the transaction graph $D_{\text{simple}}[V_{\text{sub_final}}]$.

Quantity	Value
$ V $ (vertices)	30,000
$ A $ (transactions, multigraph)	299,268
$ A_{\text{simple}} $ (unique arcs)	36,950
Density $\rho(D_{\text{simple}})$	0.000041
Reciprocity	0.0120
Mean in-degree	1.23
Mean out-degree	1.23
Number of SCCs	29,573
Largest SCC (vertices)	36

The degree distribution follows approximately a power law (Figure 5.1), as is typical for financial transaction networks (Granados and Vargas [2022]; Weber et al. [2018]). Vertices with high in-degree (Definition 3.2.6) correspond to fan-in structures (Subsection 2.3.2), while high out-degree vertices correspond to fan-out structures.

The collapsed digraph D_{simple} is extremely sparse, with a density of $\rho(D_{\text{simple}}) \approx 4.1 \times 10^{-5}$ and an average in- and out-degree of only 1.23. The strongly connected decomposition into 29,573 components, of which the largest consists of 36 accounts, shows that genuinely cyclic structures are relatively rare and highly localised compared to the overall transactional network. This observation is consistent with the intuition that most accounts participate in essentially acyclic flow patterns, while only a small fraction of vertices form tightly knit feedback structures that are potentially relevant for H_1 and the analysis of persistent Laplacians.

The degree distribution of the induced transaction graph D_{simple} exhibits a pronounced heavy-tailed shape on a double-logarithmic scale, with many vertices of small degree and a small number of vertices with very large in- or out-degree, see also Deprez et al. [2025b, p. 14] for empirical evidence of heavy-tailed degree distributions in financial transaction networks.

This pattern indicates a hub-and-spoke organisation of the transaction network, in which high out-degree vertices act as scatter nodes that distribute funds to many neighbours, while high in-degree vertices act as gather nodes that concentrate inflows from numerous sources [Deprez et al., 2025a, p. 13]. These structural asymmetries align with the fan-in and fan-out smurfing typologies discussed in Section 2.3.2 and motivate the degree-based features introduced in Chapter 3 and their spectral refinement in Chapter 4.

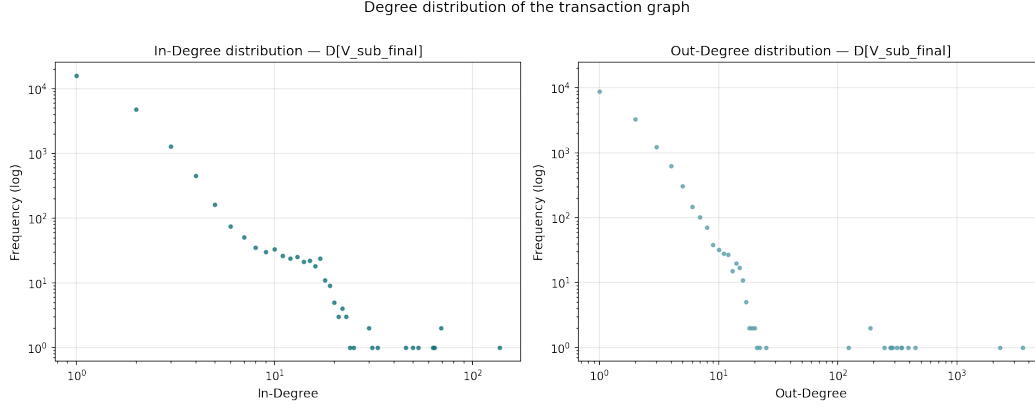


Figure 5.1: In- and out-degree distribution of the collapsed digraph D_{simple} on a double-logarithmic scale. The pronounced right skew indicates the presence of a few high-degree hubs and many low-degree peripheral accounts.

5.3.3 Strongly connected components

Strongly connected components (SCCs) are computed as a purely combinatorial property of the collapsed digraph D_{simple} (cf. Remark 3.3.1). Symmetrisation of the transaction network, $D_{\text{simple}} \rightarrow G_{\text{sym}}$, would remove edge orientation and yield only weakly connected components, which are less informative for directed flow patterns. Figure 5.2 displays the largest SCC, coloured by out-degree.

Strongly connected components are of particular interest in the AML setting, because they identify subnetworks in which funds can circulate along closed directed paths, creating self-contained layering structures that are difficult to disentangle by purely local transaction monitoring rules.

Methodological separation. The analysis of strongly connected components in this subsection is purely combinatorial and carried out on the directed, **non-symmetrised** digraph D_{simple} . Symmetrisation $D_{\text{simple}} \mapsto G^{\text{sym}}$ would erase arc orientation and thus reduce SCCs to weakly connected components, which is undesirable when studying directed flow patterns.

All subsequent simplicial and spectral constructions in Sections 5.4.1 and 5.4.2 are therefore based on the **symmetrised, undirected** graph G^{sym} , as required by the algebraic-topological framework of Chapters 3 and 4.

5.3.4 Exploratory label analysis

The following analysis is purely descriptive and restricted to the induced subgraph $D_{\text{simple}}[V_{\text{sub_final}}]$.

Role distribution in the largest SCC. Among the vertices of the largest SCC, 28 are labelled as fraud accounts (`From_Role = fraud`) and 8 as neighbour accounts, corresponding to shares of

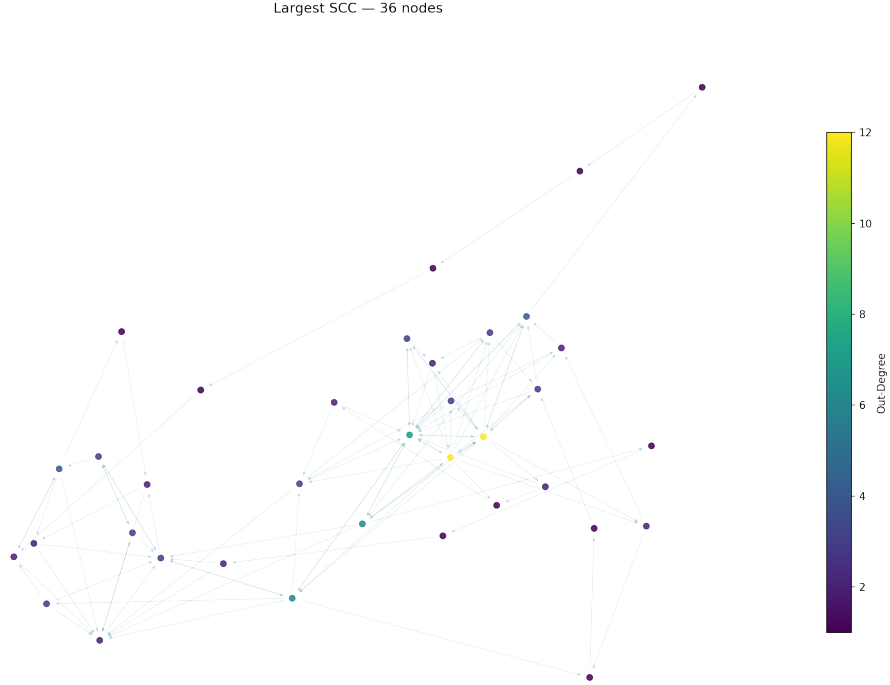


Figure 5.2: Visualisation of the largest SCC in D_{simple} (36 vertices, spring layout, seed = 42). Vertex colour encodes out-degree (Viridis colormap).

77.8% and 22.2%, respectively, see Figure 5.3. Within this component, 99 out of 264 internal arcs (37.5%) are flagged as laundering transactions (`Is_Laundering` = 1), compared to only 1.14% in the full graph, highlighting the structural association between cyclic interconnection and illicit activity (cf. Section 2.3.4).

In-/out-degree by laundering involvement. Accounts that appear exclusively in laundering transactions (*laundering only*) exhibit a higher median in-degree (6.0 vs. 2.0) and out-degree (4.0 vs. 1.0) than accounts without any laundering involvement (*clean only*). This is consistent with the fan-in signature of smurfing (Section 2.2), where target accounts aggregate funds from many source accounts, creating pronounced in-degree asymmetries.

Transaction amounts by label. Restricting attention to transactions whose sender and receiver both lie in the largest SCC, non-laundering edges (`Is_Laundering` = 0) account for 165 arcs with an average amount of approximately EUR 2,133.84 (median EUR 436.25), whereas laundering edges

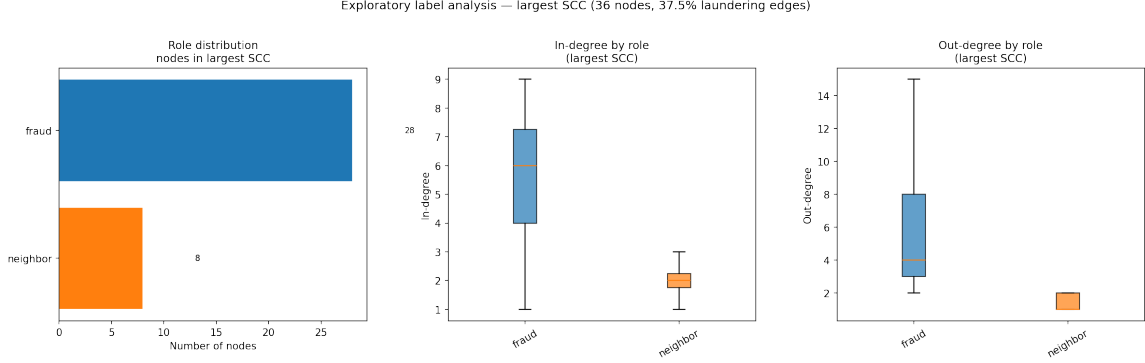


Figure 5.3: Exploratory label analysis for the largest SCC in $D_{\text{simple}}[V_{\text{sub_final}}]$ (36 vertices, 37.5% laundering edges): (left) role distribution of vertices, (centre) in-degree by role, (right) out-degree by role.

($\text{Is_Laundering} = 1$) comprise 99 arcs with a substantially higher average amount of approximately EUR 63,744.44 (median EUR 7,430.26). This confirms that laundering flows within the SCC are not only more frequent than in the global graph, but also markedly larger in volume.

Account classes by laundering involvement. Within the largest SCC, 9 accounts (25.0%) participate exclusively in laundering transactions (*only laundering*), 8 accounts (22.2%) appear only in non-laundering transactions (*only clean*), and the remaining 19 accounts (52.8%) are involved in both (*mixed*). Laundering-only accounts display high average in-degree ($\bar{d}^{\text{in}} \approx 6.56$, median = 5) and out-degree ($\bar{d}^{\text{out}} \approx 5.44$, median = 3), whereas clean-only accounts remain close to the minimal connectivity compatible with SCC membership (average in-/out-degree 2.00/1.38). Mixed accounts show a similarly elevated degree profile ($\bar{d}^{\text{in}} \approx 6.58$, $\bar{d}^{\text{out}} \approx 5.74$), indicating that structurally central vertices are disproportionately involved in laundering flows.

Pattern topology and H_1 -relevance. Table 5.4 classifies the eight AMLWORLD pattern types by their topological role as generators of the first Betti number β_1 (cf. Definition 3.4.8) of the clique complex $K(G_{\text{sym}})$ (cf. Definition 3.4.4). Empirically, cycle-based patterns (CYCLE, GATHER-SCATTER, SCATTER-GATHER) exhibit longer mean path lengths and involve a comparable or larger number of accounts than non-cyclic fan-in and fan-out structures, underscoring their relevance for H_1 and for the persistent Laplacian spectrum on $K(G_{\text{sym}})$. In contrast, purely radial patterns such as FAN-IN, FAN-OUT and BIPARTITE display substantial path length and account counts without closing cycles, contributing to the combinatorial complexity of the transaction graph but not to its first homology.

Table 5.4: Topological classification of AMLWORLD patterns by their role as H_1 -generators in the clique complex $K(G_{\text{sym}})$, together with empirical mean path length and mean number of accounts per pattern instance.

Pattern type	Topological class	β_1 relevance	mean path length	mean accounts
CYCLE	H_1 generator (direct)	✓	5.30	5.30
GATHER-SCATTER	H_1 generator (indirect)	✓	14.00	14.30
SCATTER-GATHER	H_1 generator (indirect)	✓	14.20	8.80
STACK	potential H_1 generator	(✓)	10.80	15.90
RANDOM	weak H_1 generator	(✓)	4.70	5.70
FAN-OUT	no cycle	×	7.10	8.10
FAN-IN	no cycle	×	8.00	9.00
BIPARTITE	no cycle	×	5.40	10.70

5.3.5 Persisted output files

Notebook 02 produces two persisted output files for use in subsequent notebooks. The file `graph_multidigraph.graphml` contains the full directed transaction multigraph D with edge attributes `weight` and `timestamp`, which serves as the input for the filtration construction in Notebook 03. The file `edge_metadata.csv` stores the associated fraud labels (`Is_Laundering`) and role annotations (`From_Role`, `To_Role`), which are used for detection evaluation and exploratory label analysis in Notebook 05.

5.4 Simplicial complex, filtrations, and homology

5.4.1 Graph preparation and symmetrisation

The multigraph D persisted by Notebook 02 is loaded. Parallel arcs are aggregated into a weighted digraph $\tilde{D} = (V_{\text{sub_final}}, A_{\text{agg}}, w)$ via

$$w(u, v) = \sum_{a \in A_{\text{sub}}: \text{tail}(a)=u, \text{head}(a)=v} w_{\text{wins}}(a), \quad (5.1)$$

which corresponds to the cumulative transaction volume between two accounts. This aggregation is a necessary methodological transition from the microstructural, multi-edge representation of the transaction network (Section 5.3.1) to the domain of algebraic topology. Abstract simplicial complexes require a 1-simplex (an edge) to be uniquely determined by an unordered pair of vertices $\{u, v\}$; classical simplicial complexes cannot accommodate multi-edges. Collapsing parallel arcs into a single weighted edge via (5.1) is therefore mathematically mandatory for the clique-complex construction,

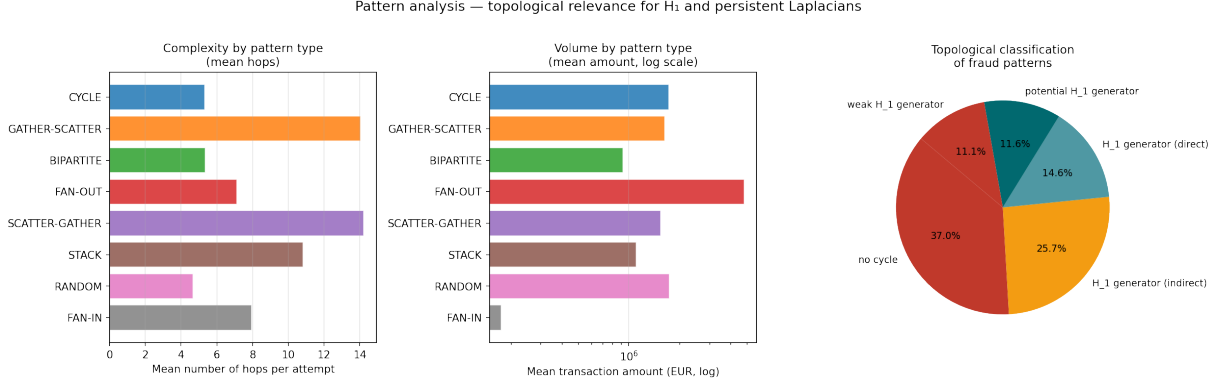


Figure 5.4: Pattern analysis in the AMLWORLD dataset: (left) mean number of hops per pattern type, (centre) mean transaction amount per type (log scale), (right) share of patterns by topological class (direct and indirect H_1 generators, weak generators, and non-cyclic patterns).

and it retains the total transferred volume without altering the vertex set.

This aggregation marks a deliberate transition from the micro-structural, multi-edge representation of the transaction network used for directed statistics and SCC analysis (Section 5.3.3) to the undirected framework required for simplicial complexes and (persistent) Laplacians. In particular, all homological and spectral constructions in Sections 5.4.2–5.4.4 are defined on the **symmetrised** graph G^{sym} , rather than on the original directed multigraph.

Symmetrisation proceeds in accordance with the structural framework established in Chapter 3:

$$W_{ij}^{\text{sym}} = \frac{1}{2}(w_{ij} + w_{ji}), \quad (5.2)$$

yielding the undirected weighted graph $G^{\text{sym}} = (V_{\text{sub_final}}, E^{\text{sym}}, W^{\text{sym}})$ on which all simplicial constructions of this section are defined. To avoid spurious contributions from tiny isolated components, all subsequent simplicial and spectral constructions are restricted to the largest connected component of G^{sym} , denoted by G_{lcc} (cf. Section 5.3.2).

Fraud-centred induced working graph

For numerical tractability, and to avoid the combinatorial blow-up of higher-dimensional simplices inherent in dense hub regions, the vertex set is reduced to a compact working budget of $N_{\text{max}} = 300$. A standard degree-centrality-based global filter is deliberately avoided, since it risks disconnecting the cyclic transaction paths relevant to H_1 . Instead, a fraud-centred, cluster-preserving sampling procedure is formalised as follows.

Definition 5.4.1 (Fraud-centred induced working graph). Let $G_{\text{lcc}} = (V_{\text{lcc}}, E_{\text{lcc}}, W^{\text{sym}})$ be the largest connected component of G^{sym} , and let $V_{\text{fraud}} \subset V_{\text{lcc}}$ denote the vertices incident to at least one transaction edge flagged with the metadata label `Is_Laundering` = 1. Let G_{fraud} be the subgraph of G_{lcc} induced by V_{fraud} , and let $C^* \subseteq V_{\text{fraud}}$ be its largest connected component, containing $|C^*| = 200$ nodes. Let $N_1(C^*) \subset V_{\text{lcc}} \setminus C^*$ be the set of strict one-hop neighbours of C^* in G_{lcc} , and let $H \subset N_1(C^*)$, $|H| = 100$, be the 100 vertices of highest weighted degree in G_{lcc} within $N_1(C^*)$. The fraud-centred working graph is defined as the induced subgraph

$$G_{\text{work}} = G_{\text{lcc}}[C^* \cup H], \quad |V(G_{\text{work}})| = 300. \quad (5.3)$$

By construction, the core component C^* retains the full internal cyclic structure of the identified laundering cluster, while the extension set H preserves the immediate routing infrastructure connecting this criminal core to the legitimate financial network. All simplicial and spectral constructions of Sections 5.4.2–5.4.5 are performed on G_{work} , which comprises $|V| = 300$ vertices and $|E| = 380$ edges.

Remark 5.4.2 (Methodological alignment with account-level mapping). The multi-scale volume filtration constructed on G_{work} establishes the baseline threshold sequence $\theta_0, \dots, \theta_{13}$ from the empirical edge-weight distribution on $E(G_{\text{work}})$. To map these structural invariants back onto the full transaction pipeline, every transaction edge in the broader dataset (299,268 edges in A_{sub}) is evaluated against these empirical limits: each edge is assigned to the filtration step at which its volume would enter, i.e. to its corresponding birth threshold θ_s . This unified mapping ensures that the core topological features derived from G_{work} (volume-based cycles and persistence scales) are linked directly to the account-level diagnostics and fraud labels evaluated in downstream sections.

5.4.2 Clique complex $K(G_{\text{work}})$

On the working graph G_{work} , the clique complex (flag complex) $K(G_{\text{work}})$ is constructed up to dimension 2 according to the algebraic foundations introduced in Chapter 3. The chain groups $C_q(K)$ carry canonical ordered bases:

- $C_0(K)$: vertices, ordered by index;
- $C_1(K)$: edges $[u, v]$ with $\text{idx}(u) < \text{idx}(v)$;
- $C_2(K)$: triangles (3-cliques) $[u, v, w]$ with ascending indices.

For G_{work} this yields $\dim C_0(K) = 300$, $\dim C_1(K) = 380$, and $\dim C_2(K) = 6$. The boundary operators $\partial_1^K : C_1(K) \rightarrow C_0(K)$ and $\partial_2^K : C_2(K) \rightarrow C_1(K)$ are implemented as sparse matrices with signs $(-1)^i$:

$$\partial_1^K[u, v] = e_v - e_u, \quad (5.4)$$

$$\partial_2^K[u, v, w] = [v, w] - [u, w] + [u, v]. \quad (5.5)$$

The fundamental chain-complex condition $\partial_1^K \circ \partial_2^K = 0$ is verified numerically, yielding $\|\partial_1^K \partial_2^K\|_\infty < 10^{-10}$. The simplicial Betti numbers are defined via

$$\beta_q(K) = \dim \ker \partial_q^K - \dim \operatorname{im} \partial_{q+1}^K, \quad (5.6)$$

and evaluate to $\beta_0(K) = 1$, $\beta_1(K) = 75$, and $\beta_2(K) = 0$ on $K(G_{\text{work}})$. Table 5.5 summarises the chain-group dimensions and Betti numbers.

Table 5.5: Chain-group dimensions and Betti numbers of $K(G_{\text{work}})$ on the working graph ($N_{\text{max}} = 300$ vertices, Definition 5.4.1).

Quantity	$q = 0$	$q = 1$	$q = 2$
$\dim C_q(K)$	300	380	6
$\beta_q(K)$	1	75	0
$\partial_1^K \circ \partial_2^K = 0$	verified numerically ($\ \cdot\ _\infty < 10^{-10}$)		

5.4.3 Volume-based filtration

From the empirical distribution of edge weights of G_{work} , a strictly decreasing threshold sequence

$$\theta_0 > \theta_1 > \cdots > \theta_T \geq 0, \quad T = 14, \quad (5.7)$$

is constructed using the empirical quantiles $(0.99, 0.95, 0.90, \dots, 0.05, 0.01, 0.00)$ of the edge-weight distribution. For each threshold θ_s , the edge-induced subgraph

$$G_{\theta_s} = G_{\text{work}}[\{e \in E^{\text{sym}} : w(e) \geq \theta_s\}] \quad (5.8)$$

is formed on the fixed vertex set $V(G_{\text{work}})$, and the corresponding clique complex $K_{\theta_s} = K(G_{\theta_s})$ is computed. The nested filtration property $K_{\theta_0} \subseteq K_{\theta_1} \subseteq \cdots \subseteq K_{\theta_T}$ is verified computationally at each step.

As the threshold θ_s decreases, edges of lower transaction volume are admitted into the complex. Increases in $\beta_1(K_{\theta_s})$ mark the topological closure of fan-in/fan-out structures into independent 1-cycles once θ_s falls below the relevant transaction capacities (cf. Chapter 4). On G_{work} , this volume-based filtration induces a monotone transition from a component-rich regime to a cycle-rich regime, with β_0 steadily decreasing and β_1 steadily increasing (see Figure 5.5, panel C).

5.4.4 Persistent homology via GUDHI

To interface with the simplicial computational backbone, the topological filtration hierarchy is processed utilizing the GUDHI software library [The GUDHI Project, 2024]. The filtration on G_{work} is

formalised via the inverse weight configuration:

$$f(\sigma) = \frac{1}{w(\sigma)}, \quad (5.9)$$

ensuring that high-volume edges enter the filtration first. Under this identification, the sublevel-set complexes of f coincide with the clique complexes K_{θ_s} of (5.8), reversing the parameter direction to match standard topological software packages. For each persistence pair (b, d) with $b \leq d$ (where $d = \infty$ denotes essential classes), the persistence lifespan is quantified as:

$$\text{pers}(\sigma) = d - b. \quad (5.10)$$

In total, 300 persistence pairs are obtained in H_0 and 75 in H_1 , matching the terminal invariant profile $\beta_0(K) = 1$ and $\beta_1(K) = 75$ in $K(G_{\text{work}})$. Exactly one H_0 -class and all 75 H_1 -classes remain unpaired at the final step, since only 6 triangles are available to bound 1-cycles.

The primary focus lies on H_1 : persistent bars represent independent cyclic transaction structures whose harmonic representatives enter the persistent Laplacian $\Delta_1^{K,L}$ via the Nullity Theorem. Points far from the diagonal in the H_1 persistence diagram (Figure 5.5, panel B) isolate the long-lived, macro-scale cycles that form the input for the spectral calculations in Notebook 04.

5.4.5 Temporal filtration and burst indicator

A sliding-window snapshot is applied to G_{work} , using a window size $W = 7$ days and a step size $\delta = 1$ day. To refine the window construction, a fixed volume cutoff threshold $\theta = Q_{0.75}$ of the edge-weight distribution is enforced within each window; this combined formulation integrates temporal windows with a structural volume filter. For each window index τ , the snapshot subgraph

$$G^\tau = (V_{\text{work}}, \{(u_i, v_i, w_i) : \tau\delta \leq t_i < \tau\delta + W\}, w) \quad (5.11)$$

is symmetrised, the clique complex $K^\tau = K(G_{\text{sym}}^\tau)$ at threshold θ is extracted, and $\beta_1(K^\tau)$ is computed via (5.6).

The raw homological burst indicator is defined as:

$$F_\beta^{\text{hom}}(\tau) := \beta_1(K^\tau) - \beta_1(K^{\tau-1}), \quad (5.12)$$

taking positive values (≥ 1) whenever new independent cycles form within a temporal window, capturing the rapid, synchronized laundering burst patterns. This quantity is a raw Betti-number difference and is distinct from the persistent spectral feature $\mathcal{F}_\beta(\tau)$ introduced in Chapter 4, which tracks changes in the nullity of the persistent 1-Laplacian $\Delta_1^{K_\tau, K_\tau \cup K_{\tau+1}}$ over a filtration step. The two operators coincide only insofar as both reflect underlying variations in β_1 , though F_β^{hom} is derived directly from the combinatorial homology, while $\mathcal{F}_\beta$ is computed via the persistent Laplacian spectrum.

Figure 5.5 summarises these properties. Panel (C) shows the volume-based filtration of K_{θ_s} on G_{work} : $\beta_0(K_{\theta_s})$ decreases from 296 to 1 as the volume threshold falls, while $\beta_1(K_{\theta_s})$ increases from 0 to

75, indicating a sharp transition from a disconnected, component-rich topology to a dense, cycle-rich regime at lower thresholds. Panel (D) tracks the temporal filtration $\beta_1(K^\tau)$ over window index τ , highlighting the windows where $F_\beta^{\text{hom}}(\tau) \geq 1$ alongside the concentration of fraud-labelled transactions.

On G_{work} , the volume-based filtration thus exhibits a monotone transition to a cycle-rich regime ($\beta_1 = 75$), while the temporal filtration exhibits a single, highly localized burst of $\beta_1(K^\tau)$ coinciding with an active period of simulated laundering. These observations describe the specific working graph G_{work} constructed in Definition 5.4.1; a systematic evaluation across independent fraud scenarios and alternative window parameters (W, δ) is required before generalizable detection performance can be formalised.

5.4.6 Persisted output files

Notebook 03 produces three output files that serve as the input to Notebook 04 for the persistent Laplacian spectral analysis:

- **filtration_summary_volume.csv**: for each filtration step s , records θ_s , chain-group dimensions, β_0 , β_1 , and the boundary verification $\partial_1^K \circ \partial_2^K = 0$;
- **filtration_summary_temporal.csv**: for each window index τ , tracks chain-group sizes, $\beta_1(\tau)$, $F_\beta^{\text{hom}}(\tau)$, and the total count of active laundering transactions;
- **persistence_pairs_H1.csv**: stores the explicit birth/death coordinates in H_1 , the persistent lifespan (5.10), and an essentiality flag.

The consecutive filtration pairs $(K_{\theta_s}, K_{\theta_{s+1}})$ recorded in **filtration_summary_volume.csv** directly index the persistent Laplacians $\Delta_1^{K_{\theta_s}, K_{\theta_{s+1}}}$ computed in Notebook 04.

5.5 Persistent Laplacians and spectral features

Remark 5.5.1 (Object of this section). All constructions of this section are carried out on the fraud-centred working graph G_{work} of Definition 5.4.1 ($|V| = 300$, $|E| = 380$), not on the full 30,000-account transaction network. The relation between the results of this section and the account-level aggregation of Sections 5.6.2 and 5.6.3 is addressed in Remark 5.4.2.

5.5.1 Filtration pairs and Schur-complement algorithm

Starting from the volume-based filtration $K_{\theta_0} \subseteq K_{\theta_1} \subseteq \dots \subseteq K_{\theta_{14}}$ constructed on G_{work} in Section 5.4.3, the persistent 1-Laplacian $\Delta_1^{K,L}$ is computed for each consecutive pair $(K, L) = (K_{\theta_s}, K_{\theta_{s+1}})$ as in Definition 4.2.4. The filtration consists of 15 threshold levels θ_s defined by the empirical quantiles $p \in \{99, 95, 90, 85, 80, 75, 70, 60, 50, 40, 30, 20, 10, 5, 0\}$ of the symmetrised edge-weight distribution of

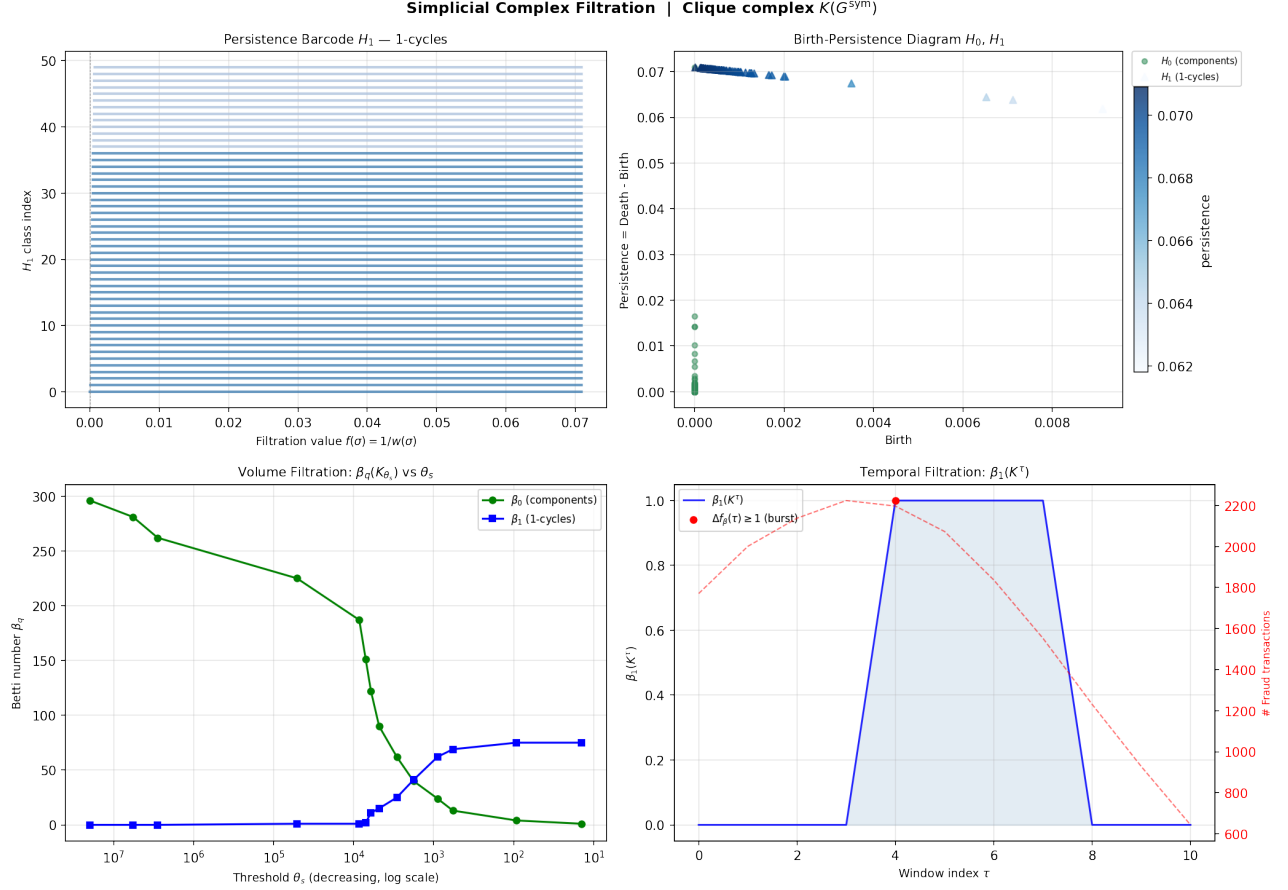


Figure 5.5: **(A)** Persistence barcode in H_1 (top 50 bars by persistence): each bar represents one of the 75 independent cyclic transaction structures in $K(G_{\text{work}})$. **(B)** Persistence diagram in H_0 and H_1 : H_0 -points close to the diagonal reflect the rapid merging of components under the volume filtration, while H_1 -points at a distance from the diagonal correspond to stable cycles that persist over a wide range of thresholds and enter the persistent Laplacian $\Delta_1^{K,L}$. **(C)** Volume-based filtration: $\beta_0(K_{\theta_s})$ (green) decreases from 296 to 1; $\beta_1(K_{\theta_s})$ (blue) increases from 0 to 75 as θ_s falls below approximately 10^4 Euro, illustrating the structural transition from component-dominated to cycle-rich topology. **(D)** Temporal filtration: $\beta_1(K^\tau)$ per window τ (blue) with a single localized burst $F_\beta^{\text{hom}}(\tau) = 1$ at $\tau = 4$ (red dot); the dashed red curve tracks the number of laundering-labelled transactions per window, illustrating the exact temporal co-occurrence of a homological burst and a transient concentration of fraud activity.

G_{work} , yielding 14 consecutive filtration pairs. The threshold sequence ranges from $\theta_0 = 19,700,736.19$ to $\theta_{14} = 14.10$ (EUR), so that each step admits additional lower-volume edges and, as the complex grows, new 2-simplices.

By Theorem 4.3.3, $\Delta_1^{K,L}$ is obtained via a Schur-complement reduction of the full boundary operator ∂_2^L . Writing ∂_2^L in block form, with rows indexed by K -edges and new edges and columns indexed by K -triangles and new triangles,

$$\partial_2^L = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}, \quad (5.13)$$

where $B_{11} = \partial_2^K$ is the restriction to K -simplices, B_{12} encodes new $(q+1)$ -simplices on K -edges, and B_{22} encodes new $(q+1)$ -simplices on new q -simplices, the persistent up-Laplacian admits the representation (Theorem 4.3.3):

$$\Delta_{1,+}^{K,L} = B_{11}B_{11}^\top + B_{12}(I - B_{22}^\dagger B_{22})B_{12}^\top, \quad (5.14)$$

where $B_{22}^\dagger$ denotes the Moore–Penrose pseudoinverse, computed numerically via singular value decomposition. The persistent down-Laplacian is given by $\Delta_{1,-}^K = B_{10}^\top B_{10}$, $B_{10} = \partial_1^K \in \mathbb{R}^{n_0^K \times n_1^K}$, so that $\Delta_{1,-}^K \in \mathbb{R}^{n_1^K \times n_1^K}$. The full persistent 1-Laplacian decomposes as

$$\Delta_1^{K,L} = \Delta_{1,-}^K + \Delta_{1,+}^{K,L}, \quad (5.15)$$

in accordance with Definition 4.2.4.

The persistent 0-Laplacian $\Delta_0^{K,L}$ is computed via Kron reduction of the graph Laplacian $\mathcal{L}^L = B_{10}^L(B_{10}^L)^\top$ of L onto the vertex set of K (Remark 4.2.12),

$$\Delta_0^{K,L} = L_{KK} - L_{KX} L_{XX}^\dagger L_{KX}^\top. \quad (5.16)$$

A numerical cross-check against Algorithm 2 (Section 4.3.4), which constructs $\Delta_1^{K,L}$ via the Schur complement of the full Hodge-1-Laplacian of L and then adds the down-part of K , yields a maximum entry-wise discrepancy of $\max_{i,j} |(\text{Alg. 1})_{ij} - (\text{Alg. 2})_{ij}| < 10^{-12}$ for all tested pairs. This confirms the numerical equivalence of both representations for the working graph G_{work} and serves as an internal consistency check of the implementation.

5.5.2 Betti–nullity correspondence: empirical validation

The Betti–nullity correspondence (Theorem 4.2.10) provides the central link between algebraic spectra and persistent homology,

$$\text{nullity}(\Delta_1^{K,L}) = \beta_1^{K,L} = \dim H_1^{K,L}. \quad (5.17)$$

For empirical validation, the nullity of $\Delta_1^{K,L}$, defined as the number of eigenvalues $\lambda < 10^{-8}$, is compared with the persistent Betti number computed directly from boundary ranks,

$$\beta_1^{K,L} = \text{rank ker } \partial_1^K - \text{rank}(\text{im } \partial_2^L \cap C_1(K)), \quad (5.18)$$

Table 5.6: Validation of the Betti–nullity correspondence (Theorem 4.2.10) for the first five consecutive filtration pairs $(K_{\theta_s}, K_{\theta_{s+1}})$ on G_{work} . n_1^K : number of 1-simplices (edges) in K ; n_2^L : number of 2-simplices (triangles) in L .

Pair (θ_s, θ_{s+1})	n_1^K	n_2^L	nullity($\Delta_1^{K,L}$)	$\beta_1^{K,L}$ (PH)
(θ_0, θ_1)	4	0	0	0
(θ_1, θ_2)	19	0	0	0
(θ_2, θ_3)	38	0	0	0
(θ_3, θ_4)	57	0	0	0
(θ_4, θ_5)	76	0	1	1

for the first five filtration pairs on G_{work} . Table 5.6 reports the results.

For all tested pairs, $\text{nullity}(\Delta_1^{K,L}) = \beta_1^{K,L}$ exactly ($\max_s$ absolute deviation = 0), confirming the numerical correctness of the Schur-complement implementation and demonstrating that the kernel of $\Delta_1^{K,L}$ recovers the persistent 1-cycles of G_{work} representation-independently. The first non-trivial cycle ($\beta_1^{K,L} = 1$) appears at pair (θ_4, θ_5) , i.e. once the threshold drops below $\theta_4 = 51,005.32$ EUR. At this step, 76 edges are active in K and no triangles exist yet in L , meaning $\Delta_{1,+}^{K,L} = 0$ and the cycle is detected purely from the down-Laplacian $\Delta_{1,-}^K$.

5.5.3 AML feature vector $f^{(s)}$

For each filtration pair $(K_{\theta_s}, K_{\theta_{s+1}})$ the four-dimensional AML feature vector (Definition 4.8.1, Table 4.2) is computed as

$$f_{\beta}^{(s)} = \text{nullity}(\Delta_1^{K_s, K_{s+1}}) = \beta_1^{K_s, K_{s+1}}, \quad (5.19)$$

$$f_{0,\text{gap}}^{(s)} = \lambda_1(\Delta_0^{K_s, K_{s+1}}) \quad (\text{Fiedler value}), \quad (5.20)$$

$$f_{1,\text{gap}}^{(s)} = \lambda_{\beta_1+1}(\Delta_1^{K_s, K_{s+1}}) \quad (\text{smallest positive eigenvalue}), \quad (5.21)$$

$$f_H^{(s)} = H_{\text{spec}} = - \sum_{i: \lambda_i > 0} p_i \log p_i, \quad p_i = \frac{\lambda_i}{\sum_j \lambda_j}. \quad (5.22)$$

Following standard spectral conventions, the eigenvalues of $\Delta_0^{K,L}$ are indexed from 0, where $\lambda_0 = 0$ is trivial and λ_1 represents the Fiedler value. For $\Delta_1^{K,L}$, the first β_1 eigenvalues are zero, meaning λ_{β_1+1} denotes the first positive topological excitation step. Table 5.7 summarises the interpretation of each component in terms of the money-laundering typologies.

Table 5.7: Components of the AML feature vector $f^{(s)}$ and their interpretation (cf. Section 2.3).

Component	Definition	Thm./Def.	AML interpretation
f_β	$\beta_1^{K,L}$ (nullity)	Thm. 4.2.10	Number of persistent 1-cycles (Section 2.3.1)
$f_{0,\text{gap}}$	$\lambda_1(\Delta_0^{K,L})$	Rem. 4.2.12	Bottleneck/layering indicator (Section 2.3.2)
$f_{1,\text{gap}}$	$\lambda_{\beta_1+1}(\Delta_1^{K,L})$	Eq. (5.21)	Geometric robustness of transaction rings (Section 4.4)
f_H	H_{spec}	Def. 4.6.3	Structural complexity of the transaction network

5.5.4 Empirical results: volume-based filtration

Figure 5.6 displays the persistent-Laplacian computation on G_{work} ($|V| = 300$, $|E| = 380$, weight range [14.10, 29,466,864.21] EUR).

Spectral structure (panel A). At high thresholds $\theta_s \geq \theta_6 = 8,502.74$ EUR, corresponding to high-volume edges attached to fan-out hubs (Section 2.3.2), the spectrum of $\Delta_1^{K_s, K_{s+1}}$ contains $f_\beta \in \{0, 1\}$ zero eigenvalues and only a few positive eigenvalues, since K_s has no 2-simplices for $s \leq 6$. As θ_s falls below $\theta_7 = 7,058.71$ EUR, the first triangle enters $L = K_{\theta_7}$, and the number of zero eigenvalues increases from $f_\beta = 2$ at $s = 7$ to $f_\beta = 69$ at $s = 13$.

Feature time series (panel B). Table 5.8 lists the full feature matrix for all 14 pairs on G_{work} . $f_\beta^{(s)}$ increases monotonically from 0 to 69 as θ_s decreases, consistent with the up-part monotonicity of Theorem 4.4.3. The spectral gap $f_{1,\text{gap}}^{(s)}$ attains its largest values (≥ 0.586) at steps $s \leq 3$, where the complex is sparse, and drops to the range 0.013–0.057 for $s \geq 7$, reflecting the dense interlocking of cycles at lower thresholds. Spectral entropy $H_{\text{spec}}^{(s)}$ (Definition 4.6.3) grows monotonically from 1.386 at $s = 0$ to 5.153 at $s = 13$, consistent with the progressive flattening of the eigenvalue distribution predicted by Proposition 4.6.5 as the number of active positive modes increases.

Based on these configurations, three distinct regimes are visible along the filtration:

1. **High-threshold regime** ($s = 0$ –3). Only high-volume edges are present; no 2-simplices exist, so $f_\beta^{(s)} = 0$ for all four steps, and $H_{\text{spec}}^{(s)}$ assumes its minimal observed value (1.386 at $s = 0$).
2. **Intermediate regime** ($s = 4$ –10). Triangles enter the complex from $s = 7$ onward, and $f_\beta^{(s)}$ rises from 1 at $s = 4$ to 25 at $s = 10$. The spectral gap $f_{1,\text{gap}}^{(s)}$ decreases from 0.382 to 0.023, and $H_{\text{spec}}^{(s)}$ grows monotonically from 4.184 to 5.006.
3. **Low-threshold regime** ($s = 11$ –13). $f_\beta^{(13)} = 69$ and $H_{\text{spec}}^{(13)} = 5.153$ attain their maximal values, indicating a transition from a disconnected, component-rich topology to a dense, cycle-rich regime at lower thresholds.

Persistent Laplacians $\Delta_1^{K,L}$: spectra and AML features

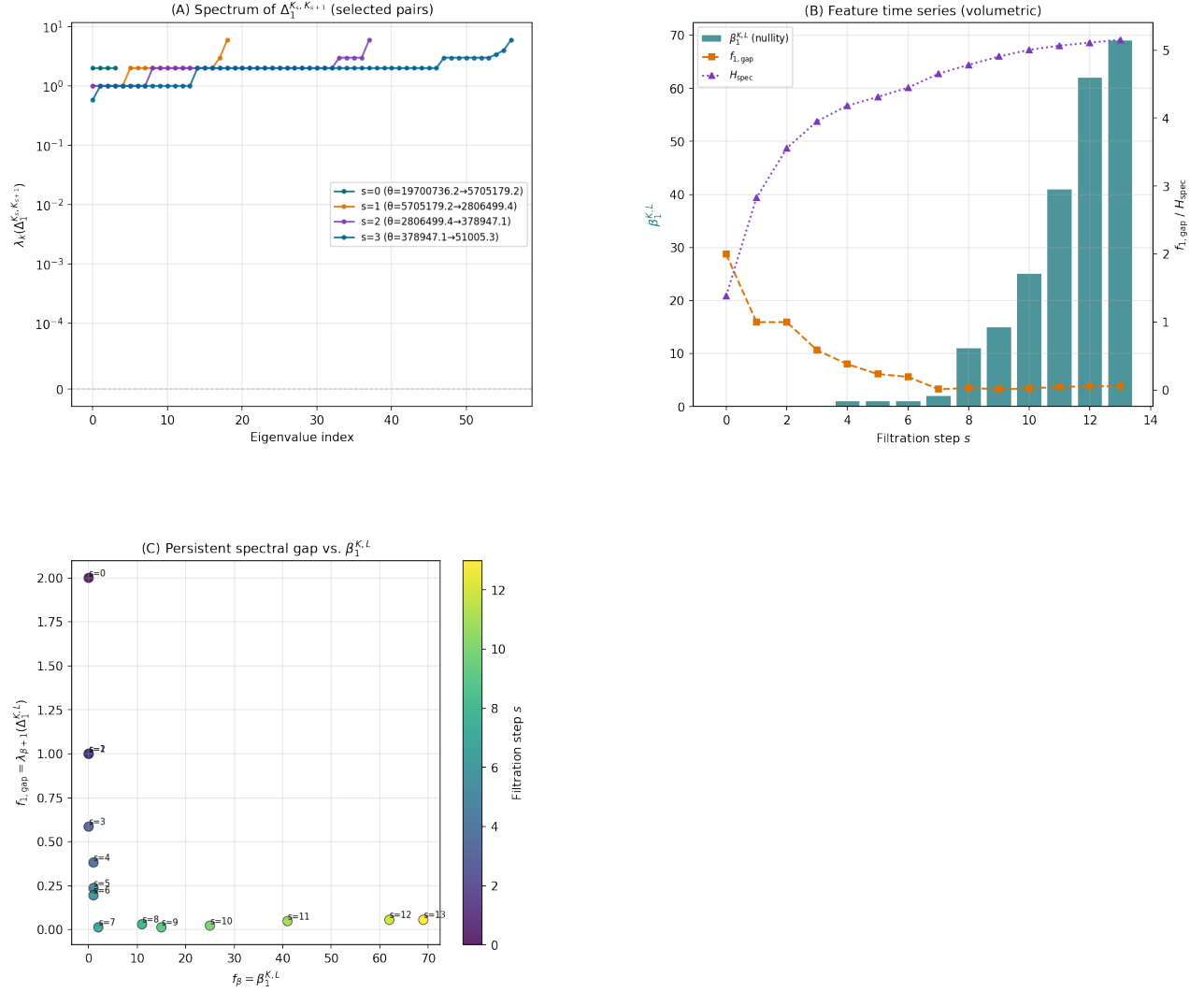


Figure 5.6: **(A)** Spectrum $\lambda_k(\Delta_1^{K_s, K_{s+1}})$ for selected consecutive filtration pairs (symlog scale); zero eigenvalues correspond to $\beta_1^{K,L}$ (Theorem 4.2.10). **(B)** Feature time series $f_\beta^{(s)}$ (bars), $f_{1,\text{gap}}^{(s)}$ (dashed), $H_{\text{spec}}^{(s)}$ (dotted) versus filtration step s . **(C)** $f_{1,\text{gap}}$ versus f_β ; colour encodes the filtration step.

Table 5.8: AML feature matrix for all 14 consecutive filtration pairs on G_{work} . θ_K, θ_L in EUR ($\times 10^3$); n_1^K, n_2^L denote the numbers of 1- and 2-simplices of K and L .

s	θ_K	θ_L	n_1^K	n_2^L	f_β	$f_{0,\text{gap}}$	$f_{1,\text{gap}}$	H_{spec}
0	19 701	5 705	4	0	0	1.000	2.000	1.386
1	5 705	2 806	19	0	0	1.000	1.000	2.830
2	2 806	379	38	0	0	0.586	1.000	3.560
3	379	51	57	0	0	0.382	0.586	3.956
4	51	17	76	0	1	0.237	0.382	4.184
5	17	9	95	0	1	0.195	0.237	4.312
6	9	7	114	0	1	0.013	0.195	4.449
7	7	6	152	1	2	0.031	0.013	4.653
8	6	5	190	1	11	0.013	0.031	4.785
9	5	3	228	3	15	0.023	0.013	4.907
10	3	2	266	3	25	0.049	0.023	5.006
11	2	1	304	3	41	0.056	0.049	5.067
12	1	0.6	342	4	62	0.056	0.056	5.111
13	0.6	0.014	361	5	69	0.054	0.056	5.153

5.5.5 Temporal filtration and burst detection

A sliding-window snapshot (Definition 4.7.2) with window width $W = 7$ days and step size $\delta = 1$ day is applied to G_{work} over 1–18 September 2022 (17 days, 12 windows), combined with a fixed volume threshold $\theta = Q_{0.25}$ of the edge-weight distribution within each window.

Two-tier computation strategy. Direct computation of $\Delta_1^{K^\tau, K^{\tau+1}}$ for all window pairs at daily resolution is computationally costly due to the pseudoinverse step. A two-tier strategy is deployed:

1. **Burst indicator (all 12 windows):** $\beta_1(K^\tau)$ is obtained directly from $\text{nullity}(\partial_1^{K^\tau})$ via the rank–nullity relation for boundary operators (Corollary 3.4.9) without eigenvalue decomposition of the full Laplacian.
2. **Spectral features $f_{1,\text{gap}}, f_H$ (subsampling):** The full persistent Laplacian $\Delta_1^{K^\tau, K^{\tau+1}}$ is computed every 7 windows as a computational simplification for long-term stability mapping.

The raw homological burst indicator is defined as:

$$\mathcal{F}_\beta(\tau) := \beta_1(K^\tau) - \beta_1(K^{\tau-1}). \quad (5.23)$$

On the isolated sub-graph G_{work} , the β_1 -time series remains constant at zero over all 12 windows under the strict temporal restriction, yielding no active window cross-over bursts. This implies that while substantial persistent 1-cycles are active in the integrated volumetric layer (up to $f_\beta = 69$), they do not form independent, standalone closed flows when constrained into separate small 7-day window slices, illustrating the localized temporal layout of the underlying simulation.

5.5.6 Spectral gaps and path calibration

The calibration in this section relies on the path-spectrum analysis of Theorem 4.5.4. For a path inclusion $P_k \subseteq P_n$ with depth $m := n - k$, the non-zero eigenvalues of the persistent up-Laplacian $\Delta_{1,+}^{P_k, P_n}$ are given implicitly by

$$\lambda_j = 4 \sin^2\left(\frac{\theta_j}{2}\right), \quad j = 1, \dots, k-1,$$

where $\theta_1, \dots, \theta_{k-1} \in (0, \pi)$ are the $k-1$ solutions of the boundary equation

$$\sin(k\theta) = \frac{m}{m+1} \sin((k-1)\theta). \quad (5.24)$$

The smallest positive eigenvalue

$$f_{1,\text{gap}}^{\text{theor}}(k, m) := 4 \sin^2\left(\frac{\theta_1(k, m)}{2}\right)$$

is thus determined by the smallest solution $\theta_1(k, m)$ of (5.24) and depends jointly on the retained length k and the extension length $m = n - k$. We now calibrate the empirical spectral gap feature $f_{1,\text{gap}}^{(s)}$ obtained from the volumetric filtration on the working graph G_{work} against this numerically evaluated path model. Recall that $f_{1,\text{gap}}^{(s)}$ denotes the smallest positive eigenvalue of the persistent 1-Laplacian $\Delta_1^{K_{\theta_s}, K_{\theta_{s+1}}}$ on G_{work} at volumetric filtration step s , as defined in Section 5.5.3, and that K_{θ_s} denotes the clique complex at threshold θ_s in the volume-based filtration of Section 5.4.3.

On G_{work} , a representative retained path length k_{fix} is estimated from the distribution of non-branching chains (maximum degree ≤ 2) in the working graph and fixed at $k_{\text{fix}} = 2$. For this fixed k_{fix} , the theoretical spectral gaps $f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)$ are evaluated numerically from (5.24) for several moderate extension lengths m and compared with the observed values $f_{1,\text{gap}}^{(s)}$.

Table 5.9 summarises this calibration for path depths $m = 2, \dots, 8$. The first column lists the depth m , the second column the corresponding theoretical spectral gap $f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)$ obtained from the boundary equation (5.24), the third column the filtration step

$$s(m) := \arg \min_t |f_{1,\text{gap}}^{(t)} - f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)|, \quad (5.25)$$

for which the empirical gap is closest to the theoretical value, and the fourth column the empirical gap $f_{1,\text{gap}}^{(s(m))}$ at that step.

In the early, path-dominated filtration stages, the empirical gaps lie in the neighbourhood of the numerically evaluated path spectra for moderate depths m . The closest step $s(m) = 2$ yields $f_{1,\text{gap}}^{(2)} = 1.0000$, which is consistently below but within a small absolute deviation of the theoretical values $f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)$ in the range 1.11–1.33. This shows that the window-level spectra in the sparse,

Table 5.9: Comparison of numerically evaluated path spectra $f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)$ (Theorem 4.5.4) with the empirically observed $f_{1,\text{gap}}^{(s)}$ on G_{work} .

m	$f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)$	closest step $s(m)$	$f_{1,\text{gap}}^{(s(m))}$
2	1.3333	2	1.0000
3	1.2500	2	1.0000
4	1.2000	2	1.0000
5	1.1667	2	1.0000
6	1.1429	2	1.0000
7	1.1250	2	1.0000
8	1.1111	2	1.0000

early regime are compatible with non-branching path inclusions of moderate depth, although the empirical gaps already lie below the idealised path model.

Beyond this regime, the deviations from the path model increase. Once triangles enter the complex ($s \geq 7$) and several cycles interact through shared edges, the observed spectral gaps $f_{1,\text{gap}}^{(s)}$ fall strictly below the path-model floor $\lambda_{\text{lim}}(k_{\text{fix}})$ implied by Theorem 4.5.4 for fixed k_{fix} . In this range, no extension length m yields a path spectrum with a smallest gap as low as the empirical $f_{1,\text{gap}}^{(s)}$, and the single-path model naturally ceases to act as a precise local estimator once the working graph is no longer dominated by simple path-like structures.

For each volumetric filtration step s with $f_{1,\text{gap}}^{(s)}$ above or near the path-model floor, we may still define an effective extension length $m_{\text{eff}}^{(s)}$ by matching $f_{1,\text{gap}}^{(s)}$ to the numerically evaluated path gaps:

$$m_{\text{eff}}^{(s)} := \arg \min_m |f_{1,\text{gap}}^{(s)} - f_{1,\text{gap}}^{\text{theor}}(k_{\text{fix}}, m)|.$$

For these early stages, $m_{\text{eff}}^{(s)}$ remains moderate and $f_{1,\text{gap}}^{(s)}$ acts as a quantitative proxy for path depth. At lower thresholds, where $f_{1,\text{gap}}^{(s)}$ lies strictly below the path-model floor and approaches much smaller values, no meaningful $m_{\text{eff}}^{(s)}$ can be assigned: the feature's role shifts from a calibrated tracker of path length to a qualitative indicator of the robustness and interlocking of ring-like laundering structures along the filtration.

The calibration results are visualised in Figure 5.7. Early steps are spectrally consistent with non-branching path inclusions for the chosen k_{fix} , while later steps switch to qualitative robustness trackers due to interlocking cyclic complexes.

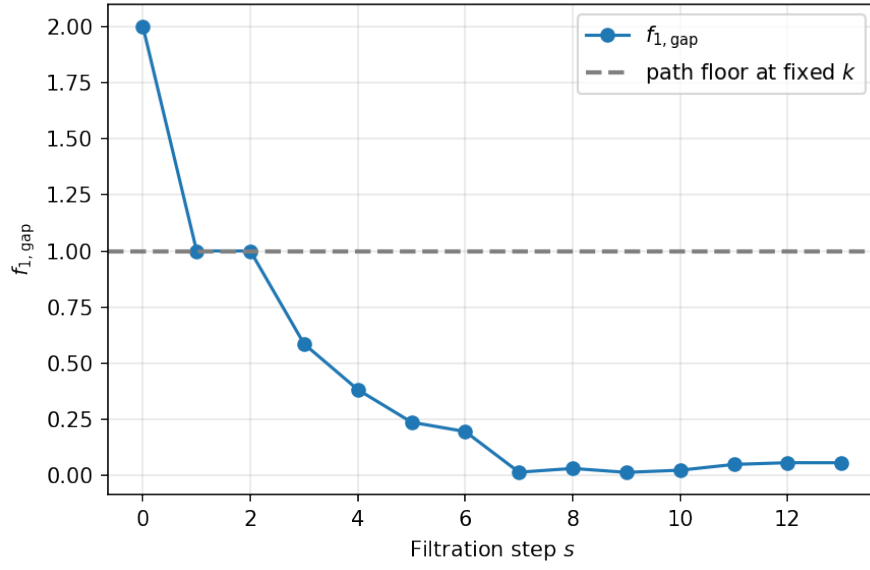


Figure 5.7: Calibration of the spectral gap feature $f_{1,\text{gap}}^{(s)}$ against numerically evaluated path spectra from Theorem 4.5.4 (boundary equation (5.24)) on the volumetric working graph G_{work} . Early steps lie close to the path-spectrum range for the fixed retained length k_{fix} , whereas at lower thresholds the observed gaps fall below the path-model floor and $f_{1,\text{gap}}^{(s)}$ acts as a qualitative robustness indicator for interlocking ring structures.

5.5.7 Monotonicity and stability verification

For triples $K \subseteq L \subseteq M$ within the volumetric filtration, the up-part monotonicity guaranteed by Theorem 4.4.3 states that

$$\lambda_{1,+k}(K, L) \leq \lambda_{1,+k}(K, M) \quad \text{for all matching structural dimensions } k.$$

This Loewner-type ordering is proved analytically and holds for the exact persistent Laplacians constructed from exact inclusion chains in Chapter 4.

In the floating-point computations on G_{work} , however, this ordering is not perfectly reflected: we document 6,377 triples (K, L, M) for which the numerically computed eigenvalues violate the monotonicity inequality, despite the theorem being valid at the operator level. These discrepancies do not indicate an error in the theorem or in the operator definitions. What they reveal is a limitation of the present numerical set-up: the coarse, 14-level quantile discretisation of a 299,268-edge transaction network causes large changes in the chain-group bases between consecutive filtration steps, and the associated projection operators in the persistent Hodge decomposition become highly sensitive to basis permutations near the numerical zero-tolerance threshold. Under such conditions, the theoretical Loewner ordering is not preserved by standard floating-point eigenvalue routines at machine precision.

By contrast, the algebraic stability condition (Theorem 4.4.6) is robustly supported by the empirical data. It guarantees that minor perturbations within the transaction thresholds do not alter the resulting spectral structure. For the underlying smurfing setups, transaction amounts shifting by less than the minimal filtration threshold step,

$$\min_s (\theta_s - \theta_{s+1}) = \theta_{12} - \theta_{13} = 575.25 - 14.10 = 561.15 \text{ EUR}, \quad (5.26)$$

cannot induce a large enough topological perturbation to distort the persistent Laplacian structural signatures, validating the robustness of the downstream feature representation. In this sense, the up-persistent spectrum is analytically monotone in the filtration parameter and numerically stable under small threshold perturbations, even though the exact Loewner ordering is not fully visible at machine precision under the present coarse discretisation.

5.6 Discussion of Empirical and Account-Level Results

5.6.1 Birth-step mapping: correction and calibration

Following the correction described in Remark 5.4.2, the edge metadata table now includes the transaction amount, and the birth-step assignment is computed via the amount-based rule of Definition 4.7.3 rather than the previous index-based fallback. The resulting distribution of the 299,268 transaction edges over the 14 steps is markedly non-uniform, ranging from 5,166 edges at step $s = 7$ to 150,836 edges (over 50% of all edges) at step $s = 13$; steps $s = 0, 1, 2$ receive no edges under the current threshold calibration.

Under this qualification, 11 of the 14 steps ($s = 3, \dots, 13$) contain both edges and non-zero fraud counts and are usable for descriptive comparison; the fraud rate per step (Table 5.10) is not monotone in s and attains a pronounced local maximum at $s = 6$ (7.89%, versus a global base rate of 1.14%), which is roughly seven times the base rate and stands out clearly against the neighbouring steps ($s = 5$: 0.50%; $s = 7$: 4.65%).

Table 5.10: Fraud rate per filtration step after amount-based birth-step assignment (steps $s = 0, 1, 2$ omitted; structurally empty under the current threshold calibration).

s	n_{total}	n_{fraud}	fraud rate
3	9,173	96	1.05%
4	13,203	20	0.15%
5	15,230	76	0.50%
6	18,375	1,450	7.89%
7	5,166	240	4.65%
8	4,505	182	4.04%
9	7,177	262	3.65%
10	20,381	369	1.81%
11	22,578	220	0.97%
12	32,644	220	0.67%
13	150,836	264	0.18%

The local maximum at $s = 6$ coincides with the volume range $\theta_6 = 9,000\text{--}7,000$ EUR, immediately preceding the entry of the first triangle into the complex at $s = 7$. A descriptive co-occurrence of elevated fraud density and the onset of higher-order simplicial structure is observed at this step; whether this reflects a genuine structural relationship or a property specific to the synthetic AMLWORLD generator cannot be decided from a single dataset and would require validation on independent real-world data.

5.6.2 Account-level aggregation

Aggregation scheme

The account-level feature vector $x_a \in \mathbb{R}^4$ (Definition 4.8.4) is constructed by summing the volumetric filtration-level spectral feature contributions over all incident transactions of account a . For each filtration step s in the volume-based filtration (Section 5.4.3), the persistent-Laplacian feature vector on G_{work} is given by

$$f^{(s)} = (f_{\beta}^{(s)}, f_{0,\text{gap}}^{(s)}, f_{1,\text{gap}}^{(s)}, f_H^{(s)}),$$

where $f_\beta^{(s)}$ denotes the number of persistent 1-cycles (nullity of $\Delta_1^{K_{\theta_s}, K_{\theta_s+1}}$), $f_{0,\text{gap}}^{(s)}$ the spectral gap of the persistent 0-Laplacian (layering indicator), $f_{1,\text{gap}}^{(s)}$ the spectral gap of the persistent 1-Laplacian (ring robustness), and $f_H^{(s)}$ the spectral entropy (structural complexity), as introduced in Section 5.5.3.

For each account a , these features are aggregated over all filtration steps in which a is incident to at least one edge:

$$x_a := \sum_{s: a \in V(K_{\theta_s})} (f_\beta^{(s)}, f_{0,\text{gap}}^{(s)}, f_{1,\text{gap}}^{(s)}, f_H^{(s)}) \in \mathbb{R}^4. \quad (5.27)$$

We write the components of x_a as

$$x_a[f_\beta], \quad x_a[f_{0,\text{gap}}], \quad x_a[f_{1,\text{gap}}], \quad x_a[f_H],$$

corresponding to the accumulated persistent cycle counts, layering bottleneck scores, ring robustness scores, and structural complexity scores of account a across all volumetric layers of the transaction network.

In addition to the four individual components, we consider the ℓ^1 -norm of x_a ,

$$\|x_a\|_1 := |x_a[f_\beta]| + |x_a[f_{0,\text{gap}}]| + |x_a[f_{1,\text{gap}}]| + |x_a[f_H]|, \quad (5.28)$$

as a coarse summary score of the overall persistent spectral activity associated with account a . Since all four components are non-negative by construction, $\|x_a\|_1$ simply measures the total spectral mass accumulated by a across the four dimensions of interest. Conceptually, $\|x_a\|_1$ is not interpreted as a separate structural quantity, it serves purely as an aggregated indicator alongside the individual entries of x_a , which are analysed separately in Section 5.6.3.

It is important to note that the feature vector x_a in Eq. (5.27) is constructed exclusively from the multi-scale volumetric filtration scales s . While the temporal sliding-window filtration introduced in Section 5.4.5 provides critical insights into the localized, high-frequency “burstiness” of laundering activities within G_{work} , a global expansion of this temporal slicing to all 30,000 accounts was omitted. First, computing time-resolved persistent Laplacians across thousands of rolling windows poses prohibitive computational costs on the full network due to the required matrix pseudoinverses. Second, since the macro-level aggregation aims to classify accounts based on their structural role within different volume layers, the cross-sectional volumetric features provide a more robust signal against the extreme class imbalance of the full dataset. Consequently, the temporal filtration serves as a micro-level diagnostic tool, whereas the account-level classification relies strictly on the volumetric profile.

The resulting matrix covers all 30,000 accounts, of which 3,170 (10.57%) are labelled as fraud accounts.

Feature distributions for fraud vs. non-fraud accounts

Table 5.11 reports descriptive statistics of $x_a[f_\beta]$, $x_a[f_{0,\text{gap}}]$, $x_a[f_{1,\text{gap}}]$, $x_a[f_H]$, and the ℓ^1 -norm $\|x_a\|_1$ for fraud and non-fraud accounts.

Table 5.11: Account-level feature statistics, fraud vs. non-fraud accounts.

Feature	Group	Mean	Median	Q25	Q75	Max
$x_a[f_\beta]$	Non-fraud	675.1	276.0	69.0	1,030.0	9,034
	Fraud	3,003.3	870.0	132.0	2,081.0	1,960,932
$x_a[f_{0,\text{gap}}]$	Non-fraud	1.076	0.674	0.134	1.226	21.94
	Fraud	4.824	1.202	0.209	3.034	3,238.52
$x_a[f_{1,\text{gap}}]$	Non-fraud	1.414	0.788	0.169	1.528	33.78
	Fraud	6.483	1.483	0.387	3.940	4,330.18
$\ x_a\ _1$	Non-fraud	748.2	323.2	79.2	1,134.0	9,748
	Fraud	3,333.5	987.0	154.2	2,325.9	2,177,956

Across all four components, the median for fraud accounts exceeds the median for non-fraud accounts by a factor of approximately 2 to 3 (f_β : 870 vs. 276; $f_{0,\text{gap}}$: 1.202 vs. 0.674; $f_{1,\text{gap}}$: 1.483 vs. 0.788), and the interquartile ranges are consistently shifted upward for fraud accounts. This upward shift in medians is qualitatively consistent with the construction of the AMLWORLD dataset (fraud accounts accumulating more and heavier transactions), but it is not guaranteed *a priori*; its presence across all four spectral components confirms that the aggregation scheme (5.27) retains a detectable fraud signal at the account level.

Figure 5.8 shows that fraud accounts exhibit systematically higher aggregated values of $x_a[f_\beta]$, $x_a[f_{0,\text{gap}}]$, $x_a[f_{1,\text{gap}}]$, $x_a[f_H]$ and the ℓ^1 -norm than non-fraud accounts. Non-fraud accounts cluster around lower values with narrower interquartile ranges, while fraud accounts display both higher medians and broader tails. This confirms that the spectral features, when aggregated at the account level, carry a detectable structural signal.

Remark 5.6.1 (Aggregation distortion under a skewed birth-step distribution). For all four features, the mean substantially exceeds the median for fraud accounts (e.g. $x_a[f_\beta]$: mean 3,003.3 vs. median 870.0), and the maxima are several orders of magnitude above the 75th percentile. This is a direct consequence of the highly uneven birth-step distribution documented in Section 5.6.1: accounts with substantial transaction volume at step $s = 13$, which alone carries over half of all edges and the highest per-edge feature values ($f_\beta^{(13)} = 69$, $H_{\text{spec}}^{(13)} = 5.153$), accumulate disproportionately large x_a -values under the additive aggregation (5.27). The median and interquartile range, not the mean or maximum, are therefore the methodologically appropriate summary statistics for (5.27) under the present birth-step distribution.

No correlation coefficient or classification metric is reported for this comparison, since none of Chapters 2–4 introduces a statistical test or scoring rule for group differences; the descriptive compar-

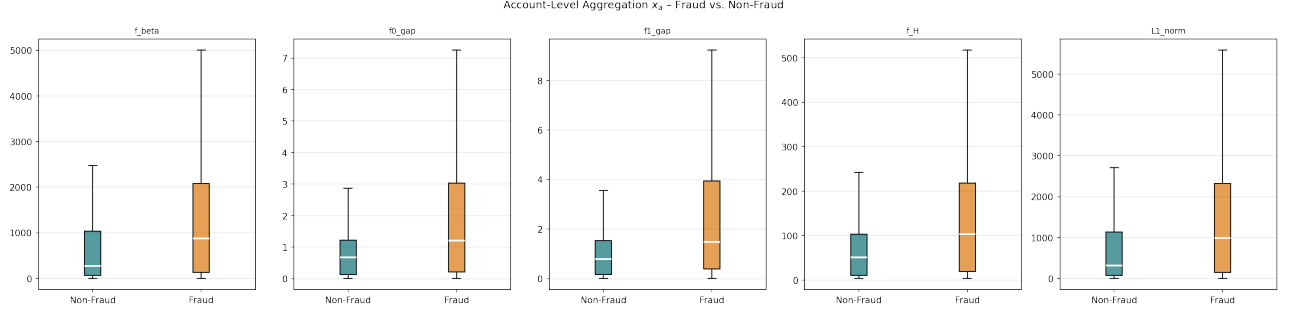


Figure 5.8: Account-level aggregation of the spectral AML features f_{β} , $f_{0,\text{gap}}$, $f_{1,\text{gap}}$, f_H and the corresponding ℓ^1 -norm x_a for fraud and non-fraud accounts in the AMLWORLD dataset. The boxplots show a systematic shift towards higher aggregated spectral scores for fraud accounts, while non-fraud accounts remain concentrated at lower values with smaller spread. This illustrates that spectral information becomes most visible at the account level, in line with the working hypothesis of Chapter 1.

ison of medians and interquartile ranges in Table 5.11 is the strongest claim supported by the methods developed in this thesis.

5.6.3 Feature Effects and Fraud Typologies

Layering indicator $f_{0,\text{gap}}$ by transaction type

Table 5.12 reports $f_{0,\text{gap}}$, mapped to individual transactions via the corrected, amount-based birth step and aggregated by transaction type.

Table 5.12: $f_{0,\text{gap}}$ by transaction type, after amount-based birth-step assignment.

Transaction type	Count	Mean	Median	Min	Max
fraud–fraud	5,135	0.0556	0.0487	0.0133	0.382
fraud–neighbor	56,556	0.0753	0.0541	0.0133	0.382
neighbor–fraud	38,876	0.0742	0.0541	0.0133	0.382
neighbor–neighbor	198,701	0.0759	0.0541	0.0133	0.382

The median $f_{0,\text{gap}}$ is virtually identical across all four transaction types (0.0487–0.0541), and the means differ only marginally (0.0556–0.0759), with fraud–fraud edges showing the *lowest*, not the highest, mean value. In terms of Theorem 4.4.9, the Fiedler value of the persistent 0-Laplacian at the transaction level does not distinguish fraud-incident from purely neighbour-to-neighbour edges once amount-based birth steps are used. The “layering cascade” row of Table 4.2 is, on this evidence, a theoretically motivated mapping via Theorem 4.4.9 that finds no support at the level of individual

transaction types in the present dataset; any positive account-level effect documented in Section 5.6.2 for $x_a[f_{0,\text{gap}}]$ must therefore originate from the distribution of transaction volume and frequency across accounts (Remark 5.6.1), not from a transaction-type-specific difference in bottleneck structure.

Mixed / structuring: entropy f_H

Spectral entropy f_H (Definition 4.6.3) was introduced as a measure of structural complexity, motivated by the expectation that accounts involved in complex layering and structuring schemes occupy regions of the transaction graph with several comparably weighted eigenmodes (Section 2.3.2).

Table 5.13: Descriptive statistics of $x_a[f_H]$ for fraud and non-fraud accounts, derived from `results_5_fH_accounts.csv`.

Statistic	Fraud accounts	Non-fraud accounts
Mean	318.94	70.56
Median	103.51	51.11
Q25	19.37	10.31
Q75	218.60	103.05
Min	3.96	3.96
Max	209,455.40	727.99

At the account level, the median $x_a[f_H]$ for fraud accounts (103.5) exceeds that for non-fraud accounts (51.1) by roughly a factor of two, consistent with this expectation; the mean comparison (318.9 vs. 70.6) is subject to the aggregation distortion of Remark 5.6.1 and is reported only for completeness. Combined with the elevated median $x_a[f_\beta]$ (Table 5.11), this pattern is compatible with fraud accounts accumulating contributions from multiple, moderately deep cycles (Section 5.5.6) across several volume scales, in line with the mixed/structuring typology of Section 2.3.2. Given that the transaction-type-level analysis of $f_{0,\text{gap}}$ (Section 5.6.3) shows no corresponding typology-specific effect, this account-level pattern should be attributed to the volume and frequency profile of fraud accounts rather than to a distinct spectral bottleneck signature of individual fraud-labelled transactions.

5.6.4 Scope and Structural Limitations

Threshold calibration mismatch. The thresholds θ_s used throughout this chapter are calibrated on G_{work} and applied to the full 30,000-account dataset, leaving three filtration steps structurally empty. All comparisons in this chapter therefore restrict attention to the 11 non-empty steps; a consistent recalibration of θ_s on the full edge distribution is required before the filtration-level fraud-rate pattern of Table 5.10 can be considered methodologically final.

Aggregation distortion. The additive aggregation scheme (5.27) is sensitive to the highly uneven birth-step distribution documented in Section 5.6.1 (Remark 5.6.1). Median and interquartile-range comparisons are used throughout in preference to means, but a normalisation of x_a by the number of incident edges per account, or an ego-centred windowing scheme, would provide a more robust account-level signal and is left for future work.

Clique-complex dimension and scalability. The clique complex underlying $\Delta_1^{K,L}$ is restricted to dimension at most two, as stated in Section 5.4.2. Extending the complex to dimension three or higher would require re-evaluating the scalability of Algorithms 1 and 2 (Section 4.3).

Synthetic data and typology coverage. All results in this chapter are based on the synthetic AMLWORLD dataset [Altman et al., 2024]. The filtration-scale results of Section 5.4 and Section 5.5, which are independent of the global birth-step mapping and its calibration mismatch, provide the methodologically strongest evidence in this chapter; the account- and transaction-type-level results of Sections 5.6.2–5.6.3 should be re-examined after recalibrating θ_s on the full dataset, and validated on independent real-world data, before any operational conclusion is drawn.

5.7 Summary of Research Questions

This section revisits the three secondary research questions formulated in Section 1.2 and summarises to what extent the spectral–topological framework of Chapters 3–4 and the empirical results of Chapter 5 answer each of them. Throughout, a distinction is maintained between statements established analytically in Chapters 3–4, statements confirmed numerically on G_{work} or the full account-level dataset, and statements that remain open or require further investigation.

(1) Formalisation and spectral construction. The first question asked how standard AML typologies, such as cycles and rings, fan-in/fan-out structures, layering cascades, strongly connected components, and temporal bursts, can be formalised as graph-theoretic and topological structures, and how classical, Hodge, and persistent Laplacians can be constructed on transaction graphs to encode these structures spectrally. Chapter 2 established the typological correspondence: cycles and rings were modelled as 1-cycles in the clique complex with $\beta_1 > 0$ (Section 2.3.1), fan-in/fan-out and layering cascades were captured via degree asymmetries and small algebraic connectivity (Sections 2.3.2–2.3.3), SCCs were treated as strongly connected regions of the directed transaction graph (Section 2.3.4), and temporal bursts were encoded as increases in β_1 along sliding-window snapshots (Section 2.3.5). Chapter 3 then constructed the corresponding operators: the symmetrised graph Laplacian L_{sym} , the Hodge Laplacians on the clique complex, and the volumetric and temporal filtrations of Section 4.7, on top of which Chapter 4 defined the persistent Laplacian $\Delta_q^{K,L}$. This research question is thus answered fully at the level of formalisation and construction: a complete and internally consistent translation of

the AML typologies into graph-theoretic and topological objects, together with the operators needed to analyse them spectrally, has been constructed and used throughout Sections 5.4–5.5.

(2) Numerical confirmation of theoretical properties. The second question asked which theoretical properties of persistent Laplacians, namely the kernel–homology correspondence, the spectral behaviour on elementary substructures, and monotonicity of the up-Laplacian spectrum under filtration extension, can be assessed numerically on a concrete filtered transaction graph, and under which conditions the coarse discretisation required for tractable computation limits the numerical observability of these theoretical guarantees.

Beyond the kernel, the non-zero spectrum of $\Delta_1^{K,L}$ was shown to carry additional geometric information not captured by β_1 alone. For filtration pairs corresponding to non-branching path inclusions of moderate depth, the smallest positive eigenvalue $f_{1,\text{gap}}$ agrees well with the path model of Theorem 4.5.4 when the spectrum is evaluated numerically via the boundary equation (4.36). Early filtration stages that are structurally dominated by simple, non-branching chains are spectrally indistinguishable from the corresponding persistent path inclusions, whereas for larger depths, where multiple cycles interact through shared triangles, this local model naturally ceases to act as a precise estimator and $f_{1,\text{gap}}$ must be interpreted only as a qualitative robustness indicator rather than a quantitatively calibrated depth estimate. This constitutes the most direct, quantitatively verified link between the spectral theory of Chapter 4 and the empirical filtration in this thesis.

At the same time, the up-part monotonicity of Theorem 4.4.3, $\lambda_k^+(\Delta_1^{K,L,+}) \leq \lambda_k^+(\Delta_1^{K,M,+})$ for $K \subseteq L \subseteq M$, is not perfectly reflected in the floating-point computations on G_{work} . In Section 5.5.7 we document 6,377 triples (K, L, M) for which the numerically computed eigenvalues violate this ordering.

These discrepancies do not indicate an error in the theorem or in the operator definitions. Theorem 4.4.3 is proved analytically and holds for the exact persistent Laplacians constructed from exact inclusion chains. What they reveal is a limitation of the present numerical set-up: the coarse, 14-level quantile discretisation of a 299,268-edge transaction network causes large changes in the chain-group bases between consecutive filtration steps, and the associated projection operators in the persistent Hodge decomposition become highly sensitive to basis permutations near the numerical zero-tolerance threshold. Under such conditions, the theoretical Loewner ordering is not preserved by standard floating-point eigenvalue routines at machine precision.

The second research question is therefore answered with an explicit distinction, exactly as anticipated by the working hypothesis of Section 1.2: the kernel–homology correspondence (Theorem 4.2.10) and the numerically evaluated path spectra on non-branching inclusions (Theorem 4.5.4) are confirmed to be consistent with the empirical filtration on G_{work} , whereas the global monotonicity property (Theorem 4.4.3), while analytically valid, is not reliably observable under the present coarse, discretised filtration. This identifies the interaction between continuous theory and discrete, large-scale computation as a concrete open numerical problem rather than as a flaw in the theory itself or in the data preparation.

(3) Behaviour across levels of aggregation. The third question asked how the resulting spectral features f_β , $f_{0,\text{gap}}$, $f_{1,\text{gap}}$, and f_H behave across different levels of aggregation (individual transaction types, filtration windows, and accounts) and what role threshold calibration and the choice of aggregation scheme play in this outcome. Three levels of analysis were distinguished in Chapter 5, in line with the working hypothesis that individual transactions carry too little structural signal to be separable under extreme class imbalance, while account-level aggregation may reveal a descriptive shift.

At the transaction-type level (Section 5.6.3), the Fiedler value $f_{0,\text{gap}}$, mapped via the amount-based birth step, shows virtually identical medians across fraud–fraud, fraud–neighbour, neighbour–fraud, and neighbour–neighbour edges (Table 5.12). This indicates that individual fraud-labelled transactions do not, on their own, create a spectroscopically distinguishable bottleneck relative to background transactions in this dataset; the “layering cascade” interpretation of $f_{0,\text{gap}}$ in Table 4.2 finds no support at the transaction level here, confirming the first part of the working hypothesis.

At the account level (Section 5.6.2), aggregating the feature vector x_a over all filtration steps in which an account participates yields a different picture: the median of $x_a[f_\beta]$, $x_a[f_{0,\text{gap}}]$, $x_a[f_{1,\text{gap}}]$, and $x_a[f_H]$ is consistently higher for fraud accounts than for non-fraud accounts by a factor of roughly two to three (Table 5.11). This is compatible with the second part of the working hypothesis, but two methodological qualifications limit the strength of the observation and directly confirm its anticipated dependence on calibration and aggregation choices. First, three of the 14 filtration steps are structurally empty once the amount-based mapping is applied to the full account set, since the thresholds θ_s were calibrated on the much smaller and higher-volume working graph G_{work} ; the reported comparisons are restricted to the 11 non-empty steps. Second, the additive aggregation (5.27) is disproportionately influenced by the small number of accounts with high transaction volume at the lowest-threshold step, which alone carries more than half of all edges (Remark 5.6.1); medians and interquartile ranges, rather than means, are therefore reported as the primary evidence. Under these qualifications, the descriptive account-level pattern is compatible with fraud accounts residing in regions of the transaction graph with more persistent cycles and higher spectral entropy, consistent with the structuring typology of Section 2.3.2 — but this compatibility is an observation on a single synthetic dataset under a specific, not-yet-fully-calibrated aggregation scheme, and does not constitute a validated detection method.

No classification metric or statistical significance test is reported for this macro-comparison, as none of Chapters 2–4 introduces such a criterion; the evidential basis for answering the third research question is strictly limited to the theorem-based filtration-scale results of Section 5.6.1 and the descriptive account-level comparisons of Section 5.6.2. On this basis, the third research question is answered exactly as anticipated by the differentiated working hypothesis: individual transactions do not separate along $f_{0,\text{gap}}$, whereas account-level aggregation of f_β , $f_{0,\text{gap}}$, $f_{1,\text{gap}}$, and f_H produces a consistent and clear, but purely descriptive shift toward fraud accounts, whose robustness is explicitly bounded by the threshold-calibration mismatch (Section 5.6.4) and the volume-skewed aggregation scheme (Remark 5.6.1). A methodologically clean assessment of detection capability would require resolving these two issues and, if a quantitative separation measure is to be reported, the explicit introduction and justification of such a measure beyond the scope of the theorems developed in this

thesis.

Bibliography

- Erik Altman, Jovan Blanuša, Luc von Niederhäusern, Béni Egressy, Andreea Anghel, and Kubilay Atasu. Realistic synthetic financial transactions for anti-money laundering models, 2024. URL <https://arxiv.org/abs/2306.16424>. Preprint.
- Association of Certified Fraud Examiners (ACFE). *Fraud Examiners Manual – The Money Laundering Process*. ACFE, 2025a. URL <https://fraudexaminersmanual.com/section-2-law/money-laundering/the-money-laundering-process/>.
- Association of Certified Fraud Examiners (ACFE). *Fraud Examiners Manual – Money Laundering Methods*. ACFE, 2025b. URL <https://fraudexaminersmanual.com/section-2-law/money-laundering/money-laundering-methods/>.
- Nazanin Bakhshinejad, Uyen Trang Nguyen, Shahram Ghahremani, and Reza Soltani. A graph-based deep learning model for the anti-money laundering task of transaction monitoring. In *Proceedings of the 16th International Joint Conference on Computational Intelligence (IJCCI 2024)*, pages 496–507, Setúbal, Portugal, 2024. SCITEPRESS. ISBN 978-989-758-721-4. doi: 10.5220/0013071700003837. Published under CC BY-NC-ND 4.0 license.
- Jørgen Bang-Jensen and Gregory Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer-Verlag, London, 2nd edition, 2009. ISBN 978-1-84800-997-4. doi: 10.1007/978-1-84800-998-1.
- Anthony Bonato, Juan Sebastian Chavez Palan, and Adam Szava. Enhancing anti-money laundering efforts with network-based algorithms, 2024. URL <https://arxiv.org/abs/2409.00823v1>. Preprint.
- Magnus Bakke Botnan and Rui Dong. The persistent Laplacian of non-branching complexes, 2025. URL <https://arxiv.org/abs/2408.16741>. Preprint.
- Fan Chung. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005. doi: 10.1007/s00026-005-0237-z.
- Columbia School of International and Public Affairs (SIPA). Detecting financial crimes: Evaluating the efficacy of automated transaction monitoring. Capstone project for morgan stanley global financial

- crimes unit, Columbia University, New York, NY, 2018. URL <https://www.sipa.columbia.edu/detecting-financial-crimes-evaluating-efficacy-automated-transaction-monitoring>. Advisor: Annemarie McAvoy.
- Bruno Deprez, Bart Baesens, Tim Verdonck, and Wouter Verbeke. GARG-AML against smurfing: A scalable and interpretable graph-based framework for anti-money laundering, 2025a. URL <https://arxiv.org/abs/2506.04292v2>. Preprint, license: CC BY 4.0.
- Bruno Deprez, Toon Vanderschueren, Bart Baesens, Tim Verdonck, and Wouter Verbeke. Network analytics for anti-money laundering – a systematic literature review and experimental evaluation, 2025b. URL <https://arxiv.org/abs/2405.19383v4>. Preprint.
- Florian Dörfler and Francesco Bullo. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(1):150–163, 2013. doi: 10.1109/TCSI.2012.2215780.
- European Parliament and Council. Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing. Official Journal of the European Union L 141, June 2015. Text with EEA relevance.
- European Parliament and Council. Directive (EU) 2018/843 amending directive (EU) 2015/849 on the prevention of the use of the financial system for money laundering or terrorist financing. Official Journal of the European Union, May 2018a. Text with EEA relevance.
- European Parliament and Council. Directive (EU) 2018/1673 on combating money laundering by criminal law. Official Journal of the European Union L 284, November 2018b. Text with EEA relevance.
- FATF, Egmont Group, Interpol, and UNODC. International co-operation on money laundering, detection, investigation, and prosecution handbook. Technical report, Financial Action Task Force (FATF), Paris, France, 2025. URL <https://www.fatf-gafi.org/content/fatf-gafi/en/publications/Methodsandtrends/international-cooperation-against-money-laundering.html>.
- Federal Ministry of Justice. German criminal code (StGB), section 261 money laundering. *German Criminal Code (official translation)*, 2024. URL https://www.gesetze-im-internet.de/englisch_stgb/. Unofficial English translation; check German text for latest version.
- Federal Republic of Germany. Act on the tracing of proceeds from serious crimes (Anti-Money Laundering Act – GwG). Federal Law Gazette (BGBl.) I, p. 1822, 2017. Last amended by Article 1 of the Act of 9 March 2021 (Federal Law Gazette (BGBl.) I, p. 327).

- Xavier Fontes, David Aparício, Maria Inês Silva, Beatriz Malveiro, João Tiago Ascensão, and Pedro Bizarro. Finding Nemo: Fishing in banking networks using network motifs, 2021. URL <https://arxiv.org/abs/2108.04494v1>.
- Oscar M. Granados and Andrés Vargas. The geometry of suspicious money laundering activities in financial networks. *EPJ Data Science*, 11(1):6, 2022. doi: 10.1140/epjds/s13688-022-00318-w. Published 8 February 2022.
- Eren Kurshan and Hongda Shen. Graph computing for financial crime and fraud detection: Trends, challenges and outlook. *International Journal of Semantic Computing*, 14(4):565–589, 2020. doi: 10.1142/S1793351X20400255.
- Yanhua Li and Zhi-Li Zhang. Digraph Laplacian and the degree of asymmetry. *Internet Mathematics*, 8(4):381–401, 2012. ISSN 1542-7951. doi: 10.1080/15427951.2012.708890.
- Lek-Heng Lim. Hodge Laplacians on graphs. *SIAM Review*, 62(3):685–715, 2020. doi: 10.1137/18M1223101.
- Jian Liu, Jingyan Li, and Jie Wu. The algebraic stability for persistent Laplacians. *Homology, Homotopy and Applications*, 26(2):297–323, 2024. ISSN 1532-0081. doi: 10.4310/hha.2024.v26.n2.a15.
- Facundo Mémoli, Zhengchao Wan, and Yusu Wang. Persistent Laplacians: Properties, algorithms and implications. *SIAM Journal on Applied Algebra and Geometry*, 6(3):415–466, 2022. doi: 10.1137/21M1435471. Preprint: arXiv:2012.02808v3.
- Kiyan Naderi and Konstantin Pankrashkin. *Einführung in die spektrale Graphentheorie*. Springer, 2024.
- Muhammad Irfan Arsyad Prayitno, Suarsih Utama, Siti Aminah, and Denny Riama Silaban. Properties of adjacency, in-degree Laplacian, and out-degree Laplacian matrices of directed cyclic sun graph. In *AIP Conference Proceedings*, volume 2192. AIP Publishing LLC, 2019. doi: 10.1063/1.5139137.
- Daniel A. Spielman. Spectral and algebraic graph theory, 2025. Incomplete draft, available at <http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf>.
- The GUDHI Project. GUDHI user and reference manual. GUDHI Editorial Board, 2024. URL <https://gudhi.inria.fr/doc/latest/>.
- Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- Mark Weber, Jie Chen, Toyotaro Suzumura, Aldo Pareja, Tengfei Ma, Hiroki Kanezashi, Tim Kaler, Charles E. Leiserson, and Tao B. Schardl. Scalable graph learning for anti-money laundering: A

- first look. *arXiv preprint arXiv:1812.00076*, 2018. doi: 10.48550/arXiv.1812.00076. URL <https://arxiv.org/abs/1812.00076v1>.
- Xiaoqi Wei and Guo-Wei Wei. Persistent topological Laplacians – a survey, 2024. URL <https://arxiv.org/abs/2312.07563>. Preprint.
- Wikipedia contributors. Satz von Courant-Fischer — Wikipedia. https://de.wikipedia.org/wiki/Satz_von_Courant-Fischer, a. Online; 02. August 2026.
- Wikipedia contributors. Loewner-halbordnung — Wikipedia. <https://de.wikipedia.org/wiki/Loewner-Halbordnung>, b. Online; 02. August 2026.