ADAPTIVE FIDELITY-DRIVEN RECONSTRUCTION (AFR): A REALISTIC THREAT MODEL FOR SPECTRAL EMBEDDING LEAKAGE

Anonymous authors
Paper under double-blind review

ABSTRACT

The exchange of structural representations in Federated Graph Learning (FGL) creates a potent channel for privacy leakage. While theoretical graph reconstruction is possible, existing attack models are brittle as they hinge on an unrealistic assumption: perfect, noise-free local data. This paper elevates that theoretical threat to a practical reality. We introduce AFR (Adaptive Fidelity-driven Reconstruction), a robust new attack model that abandons idealized assumptions. Instead of assuming data quality, AFR actively measures and exploits it. The algorithm first quantifies the reliability of each local patch via a novel fidelity score, combining a spectral signal-to-noise ratio with structural entropy. This score then guides a robust assembly process that uses RANSAC-Procrustes to tolerate outliers and adaptive stitching criteria to manage uncertainty. Instead of a single, perfect graph, AFR recovers large, high-fidelity, and internally consistent islands from the most trustworthy data. Experiments on the LoGraB benchmark show that AFR successfully reconstructs significant topology in challenging, noisy regimes where idealized models fail completely. Our work thus promotes spectral leakage from a theoretical possibility to a practical and potent threat. Our source code is anonymously available at: `https://anonymous.4open.science/r/AFR-ICLR-submission`.

1 INTRODUCTION

Federated Learning (FL) promises a future of privacy-preserving machine learning, where models are trained collaboratively without sharing raw data. This promise rests on a core premise: that exchanging only intermediate representations, like gradients or embeddings, is safe. But this premise is fragile. The intermediate signals themselves can inadvertently leak sensitive information. In Federated Graph Learning (FGL), this challenge is amplified. The primary asset to protect is not just the node features, but the local graph topology, which can encode sensitive relationships. Recent comprehensive evaluations have confirmed the severity of this threat through various reconstruction and inference attacks (Chen et al., 2024). This vulnerability extends beyond topology, with emerging work showing leakage of sensitive node properties (Liu et al., 2025) and even local label distributions from shared representations. Spectral embeddings, a common currency for exchanging this structural information, create the most potent passive leakage channel. Unlike active attacks that risk detection, a passive attack requires only observing a single, honestly-shared artifact. The vulnerability, therefore, lies not in the federated protocol, but in the intrinsic properties of the representation itself. This threat is more fundamental, challenging core assumptions about the safety of distributed graph learning.

To understand the practical threat of spectral leakage, we first consider a best-case scenario. We construct an idealized stitching model, a baseline that proves local spectral embeddings can, in principle, be stitched together for exact global reconstruction. This confirms that simple data partitioning is not an effective defense. However, the power of this model depends on a set of unrealistic conditions. It demands a near-unattainable level of data perfection, requiring both: (1) that every local patch exhibits a significant spectral gap for noiseless inversion and (2) that every adjacent patch overlap is large and well-connected for perfect alignment. This is the "curse of perfection". The reconstruction is a deterministic cascade; a single patch that violates these conditions introduces an initial error that propagates and catastrophically corrupts the entire global assembly. The approach is thus fundamentally brittle. This brittleness reveals a critical gap between theoretical

possibility and practical threat, motivating the central question of our work: Can an adversary reconstruct a meaningful topology when the leaked information is inevitably noisy, incomplete, and of heterogeneous quality?

Our answer is **AFR** (Adaptive Fidelity-driven Reconstruction), an algorithm built on a new philosophy: Instead of assuming that the data are perfect, it embraces its imperfection. AFR actively measures, quantifies, and exploits the varying quality of leaked information. First, it assigns each patch a novel fidelity score (s_v), combining spectral stability (SNR) and topological complexity (Structural entropy) to gauge its reliability (shown in Section 3.2). Only the most trustworthy core patches are even considered for assembly. The assembly process is then designed for failure. A robust RANSAC-Procrustes alignment handles the inevitable outliers from noisy reconstructions, and an adaptive stitching threshold demands more evidence (a larger overlap) to connect less reliable patches. This fidelity-driven approach redefines success. AFR abandons the fragile pursuit of a single global graph. Instead, its goal is to recover a group of large, internally-consistent, high-fidelity islands from the most reliable data, while intelligently isolating what cannot be recovered.

Our work makes the following contributions:

- **A practical threat model for spectral leakage**. We are the first to formalize and address spectral leakage under realistic and imperfect conditions. Our approach accounts for noise, reconstruction errors, and heterogeneity, elevating the threat from a theoretical possibility to a practical concern.
- **A computable fidelity score**. We introduce a novel fidelity score that quantifies the reliability of leaked graph patches. By combining a spectral signal-to-noise ratio (SNR) with structural entropy, this score provides a quantitative foundation for robust decision-making.
- **Adaptive Fidelity-driven Reconstruction (AFR) algorithm**. We propose AFR, an adaptive, multi-stage algorithm that uses our fidelity scores to guide a robust assembly process. Key components include an outlier-resilient RANSAC-Procrustes alignment and adaptive stitching criteria to manage uncertainty.
- **Extensive empirical validation**. We conduct experiments across a wide range of realistic, challenging scenarios generated by our evaluation protocol. Our results demonstrate that AFR successfully reconstructs significant topology in regimes where idealized prior methods fail completely, thus confirming the viability of our new threat model.

2 Related work

Privacy threats in Federated Graph Learning The premise that sharing intermediate representations in Federated Learning (FL) is safe has been systematically challenged. In particular, Deep Leakage from Gradients (DLG) demonstrated that raw training data can be recovered from shared gradients (Zhu et al., 2019; Geiping et al., 2020). Although our work shares this inversion philosophy, we target a fundamentally different leakage channel: not dynamic gradients, but static spectral embeddings of the graph structure. This vulnerability applies to any FGL scenario in which structural representations are exchanged, even if no gradients are shared. Within the taxonomy of graph privacy attacks, AFR instantiates a passive eavesdropper threat model. As summarized in the table below, this is distinct from active query-based attacks such as link inference (He et al., 2021b) or membership inference (Zhang et al., 2021), which risk detection. It also differs from malicious-client attacks. The passive model represents a more fundamental threat, as the vulnerability is an intrinsic property of the honestly-shared data artifact itself, making it virtually undetectable at the protocol level. Comprehensive empirical studies have formalized Graph Reconstruction Attacks (GRA) as a major vulnerability (Chen et al., 2024). Beyond structural recovery, sophisticated influence-based attacks have been developed to infer links with high accuracy (Wu et al., 2022; Meng et al., 2023). The severity of membership inference, which AFR's threat model is related to, has prompted the recent development of dedicated defenses, highlighting the community's focus on this problem (Zhong et al., 2025).

Graph reconstruction from Local Embeddings Our work builds upon the theoretical principle that a global graph can, under idealized conditions, be perfectly reconstructed by stitching together local spectral embeddings. The fundamental feasibility of inverting spectral representations is supported by theoretical findings demonstrating that, under certain identifiability conditions, graphs

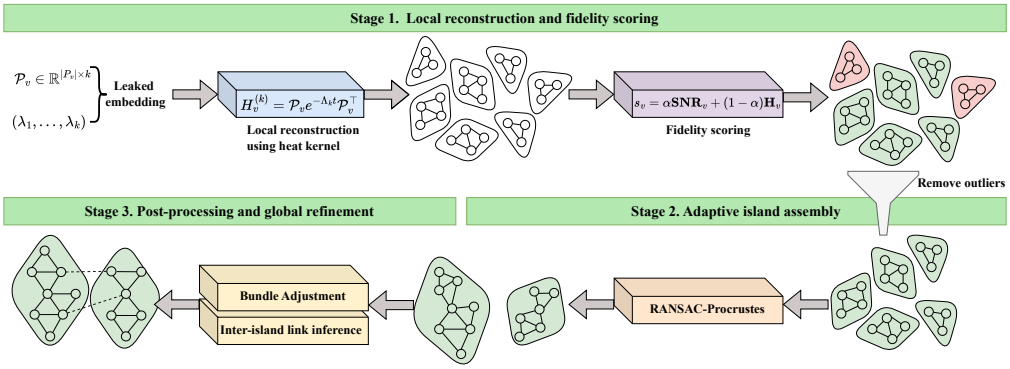

Figure 1: **Adaptive Fidelity-driven Reconstruction (AFR)**. The pipeline reconstructs graphs in three stages: local reconstruction with fidelity scoring, adaptive island assembly using RANSAC-Procrustes, and global refinement via Bundle Adjustment and inter-island link inference.

are reconstructible from their eigenspaces, even in the presence of noise (Castro et al., 2017). But this guarantee comes with a curse of perfection: it demands that every patch and every overlap be flawless. A single imperfect patch is enough to trigger a catastrophic cascade of errors, which makes the entire pipeline fundamentally brittle. AFR is designed precisely to break this curse, operating on the noisy, imperfect embeddings found in realistic settings. Conceptually, our problem can be framed as an "Inverse GAE". While a Variational Graph Auto-Encoder (Kipf & Welling, 2016) learns a mapping from a graph to an embedding, $p(\mathcal{A}|\mathcal{Z})$, our challenge is the inverse: given a collection of corrupted localized embeddings $\mathcal{P}$, we must infer a plausible graph, $p(\mathcal{A}|\mathcal{P})$.

Robust geometric alignment The resilience of AFR is inspired by battle-tested techniques from computer vision, where dealing with noisy and imperfect data is the norm. Standard Orthogonal Procrustes analysis is notoriously sensitive to outliers (Schönemann, 1966). To overcome this, we replace it with RANSAC-Procrustes (Fischler & Bolles, 1981), classic paradigm designed to find a reliable fit even in the presence of faulty data points. To correct for small alignment errors that accumulate into large-scale distortions, AFR employs a post-processing step analogous to Bundle Adjustment (Triggs et al., 2000), a cornerstone of structure-from-motion that jointly optimizes for global geometric consistency.

Defenses against information leakage in FGL The significant privacy risks highlighted by AFR and related works have spurred the development of various defense mechanisms. A prominent line of defense involves applying Differential Privacy (DP). This includes adding noise to shared parameters to provide formal privacy guarantees (Qiu et al., 2022) or perturbing the aggregation function within GNNs to protect edge-level privacy (Sajadmanesh et al., 2023). Another approach uses information-theoretic principles, such as the information bottleneck, to learn minimal, privacy-preserving subgraphs that are shared instead of the original topology (Zhang et al., 2023). Finally, hybrid methods combine techniques like Local Differential Privacy (LDP) with secure hardware such as Intel SGX to create multi-layered protection (Han et al., 2024). These defensive strategies underscore the urgent need for realistic threat models such as AFR to properly evaluate their effectiveness.

In summary, while previous work has separately explored reconstruction attacks or robust alignment, AFR is the first to synthesize these principles into a coherent framework that addresses the practical threat of leakage from imperfect, localized spectral embeddings. Our work is at the intersection of privacy in federated systems, graph reconstruction attacks, and robust geometric alignment.

3 Adaptive Fidelity-driven Reconstruction (AFR)

3.1 Theoretical underpinnings

Our theoretical framework is built upon a set of formal assumptions that guide the algorithm's behavior under realistic, imperfect conditions. Instead of demanding data perfection, we assume

that a subset of the local data is sufficiently reliable to initiate a robust reconstruction and that this reliability can be quantified. First, we formally define a core patch as a local subgraph that is computationally tractable, has a bounded reconstruction error from its spectral embedding, and crucially, surpasses a minimum threshold on our composite fidelity score. This score ensures that we only build upon patches that are both spectrally stable and structurally informative. Second, we establish an adaptive eligibility criterion for stitching, assuming that a pair of core patches can only be aligned if their overlap is sufficiently large and structurally sound. Critically, this required overlap size increases as the fidelity of the patches decreases, forcing the algorithm to demand stronger evidence when faced with greater uncertainty. Finally, we rely on the standard probabilistic guarantees of our alignment method, assuming that for any eligible pair, RANSAC-Procrustes can recover a near-correct relative rotation with high probability. These assumptions collectively create a principled foundation for a reconstruction algorithm that is resilient to noise and heterogeneity. The complete mathematical formulation of these assumptions is detailed in Appendix A.

3.2 METHODOLOGY

Let $G = (V, E)$ be an undirected graph, $|V| = n$. For every $v \in V$ we observe a local d-hop embedding $\mathcal{P}_v \in \mathbb{R}^{|P_v| \times k}$ of the patch $P_v \subseteq V$. Our goal is to reconstruct a set of high-fidelity *islands* $\{\widehat{G}_1, \ldots, \widehat{G}_r\} \subseteq G$ together with a probabilistic inter-island edge set $\widehat{E}_{\text{cross}}$.

Stage 1. Local reconstruction and fidelity scoring The first stage of AFR acts as a gatekeeper. Its goal is to transform each raw, leaked spectral embedding into a structured representation and then quantify how trustworthy it is as an anchor for reconstruction.

Local reconstruction. For each patch v, we reconstruct a local adjacency matrix $\widehat{\mathcal{A}}_v$ from its spectral embedding $\mathcal{P}_v \in \mathbb{R}^{|P_v| \times k}$ and corresponding eigenvalues $(\lambda_1, \ldots, \lambda_k)$ using the Heat kernel method. This inverts the embedding process to produce a tangible estimate of the local topology. We first form the truncated heat kernel approximation: $H_v^{(k)} = \mathcal{P}_v e^{-\Lambda_k t} \mathcal{P}_v^\top$, where t is a small fixed-time parameter and $\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$. The entries of the true heat kernel exhibit a positive separation gap between the connected and disconnected nodes, allowing for accurate recovery. The local adjacency matrix $\widehat{\mathcal{A}}_v$ is then obtained by applying a threshold τ_v to the entries of $H_v^{(k)}$.

Fidelity scoring. But a reconstruction is only as good as its source data. The core innovation of this stage is a multi-component fidelity score, s_v, designed to answer two orthogonal questions about each patch's quality:

1. **Is the signal clean?** We quantify this with the spectral **signal-to-noise ratio (SNR)**. This ratio measures the stability of the local eigenspace against truncation errors, giving us a principled measure of information quality.
$$\text{SNR}_v = \frac{\delta_v}{\delta_v + \eta_v}.$$
where the signal $\delta_v = \lambda_{k+1}(\mathcal{L}_v) - \lambda_k(\mathcal{L}_v)$ and the noise $\eta_v = e^{-t\lambda_{k+1}(\mathcal{L}_v)}$. This formulation provides the information quality relative to the information lost by truncation.
2. **Is the signal distinctive?** We quantify this with **Structural entropy**. This metric gauges the topological complexity by using the normalized Shannon entropy ($\mathcal{E}_v$) of the patch's degree distribution.
$$\mathcal{E}_v = -\frac{1}{\log\big(|V_v|\big)} \sum_{d \in D_v} p(d)\ \log\big(p(d)\big),$$
where D_v is the set of unique node degrees in the patch and $p(d)$ is the empirical probability that the node has degree d. A patch might have a clean signal but a trivial structure (low entropy) or be complex but noisy (low SNR). A trustworthy patch needs both.

These two components are then combined into a composite fidelity score via a convex combination:
$$s_v = \alpha \text{SNR}_v + (1 - \alpha)\mathcal{E}_v,$$
where $\alpha \in [0, 1]$ balances the relative importance of spectral stability and structural complexity. Finally, we filter for quality: Only patches that surpass the minimum threshold for both their fidelity score and their spectral gap are promoted to "core patches", forming the trusted input for the next stage.

Stage 2. Adaptive island assembly This stage is the heart of AFR, where the trustworthy "core patches" are intelligently assembled into larger, internally consistent structural "islands".

Data-driven prioritization. To minimize error propagation, the assembly follows a prioritized scheme. The set of core patches $(\widehat{\mathcal{A}}_v, s_v)$ that satisfy Assumption 5 forms the input to this stage. The priority queue then determines the next best potential "stitch", based on a combination of the joint fidelity of the patch pair $f(s_v, s_w)$ and the size of their overlap $|I_{vw}|$.

Adaptive stitching loop. The algorithm iteratively attempts to stitch the highest-priority pair (v, w) from the queue. This process involves a rigorous multi-step verification:

1. Eligibility check: First, the pair must satisfy our criterion: Their overlap must be large enough and the induced subgraph must be structurally connected. This overlap size threshold increases for lower-fidelity patches, forcing the algorithm to demand stronger evidence when faced with more uncertainty.
2. Robust alignment: If eligible, we align the pair using RANSAC-Procrustes. This choice is deliberate: It is designed to find a correct rotation even in the presence of faulty correspondences, an inevitability given reconstruction errors (see details in Algorithm 1).
3. Stitching decision: A stitch is accepted only if RANSAC finds a strong geometric consensus in a significant portion of the overlap (max_consensus_count $\geq d_{\text{adaptive}}(s_v, s_w)$). This final check ensures that the stitches are based on coherent evidence, making the entire assembly process highly robust.

Stage 3: Post-processing and global refinement The final stage of AFR performs two operations: refining the internal geometric consistency of each island and inferring latent connections between them.

Intra-island refinement via Bundle Adjustment. First, we correct for "drift". Even with robust pairwise stitching, small alignment errors can accumulate across large islands, causing subtle distortions. To fix this, we employ a global refinement step analogous to Bundle Adjustment from computer vision. We jointly optimize all pairwise rotations within an island to minimize a global geometric error, effectively "truing up" the entire structure.

$$\{\mathcal{Q}^*_{vw}\} = \text{argmin}_{\{\mathcal{Q}_v w\}} \sum_{(v,w)\in \text{stitches}(G_i)} \|\mathcal{P}'_v - \mathcal{P}'_w \mathcal{Q}_{vw}\|_F^2 ,$$

where $\mathcal{P}'_v$ and $\mathcal{P}'_w$ are the eigenvector submatrices corresponding to the overlap. This non-convex optimization problem over the manifold of orthogonal matrices can be solved using iterative methods such as Riemannian gradient descent (see details in Appendix E.5).

Inter-island link inference via cross-voting. Finally, we hunt for the potential links between our recovered islands. To do this, we return to the raw data and use a cross-voting mechanism. The intuition is simple: if two nodes from different islands u and w frequently co-occurred in the initial, low-quality patches, it constitutes strong latent evidence of a true connection. We formalize this by computing a vote count, $C(u, w)$, and if it surpasses a threshold, we infer a probabilistic edge. This allows AFR to reason about the global topology, even the parts it could not confidently reconstruct.

$$P_{inter}(u, w) = \frac{1}{1 + e^{-\beta(C(u,w)-C_0)}},$$

where C_0 is a vote threshold and β controls the steepness of the probability curve.

3.3 Theoretical guarantees

Notation. For a local patch P_v with $n_v = |P_v|$ nodes, let $\mathcal{L}_v$ be its (combinatorial or normalized) Laplacian with eigenpairs $\{(\lambda_r, u_r)\}_{r\geq 1}$, ordered nondecreasingly. The truncated heat kernel is

$$H_v^{(k)}(t) = \sum_{r=1}^{k} e^{-t\lambda_r} u_r u_r^\top, \quad \text{and the residual } R_v^{(k)}(t) = H_v(t) - H_v^{(k)}(t) = \sum_{r>k} e^{-t\lambda_r} u_r u_r^\top.$$

We reconstruct A_v by thresholding the entries of $H_v^{(k)}(t)$. The patch-wise spectral gap is $\delta_v = \lambda_{k+1} - \lambda_k$ and the truncation proxy is $\eta_v = \|R_v^{(k)}(t)\|_2 = e^{-t\lambda_{k+1}}$. The composite fidelity is $s_v = \alpha \cdot \text{SNR}_v + (1 - \alpha) \cdot \mathcal{E}_v$ with $\text{SNR}_v = \delta_v/(\delta_v + \eta_v)$ and $\mathcal{E}_v$ the structural entropy.

Theorem 1 (Edge recovery). *Let $t > 0$ be a fixed time parameter. Suppose the true heat kernel $H_v(t)$ on P_v exhibits an entry-wise separation margin between its edge and non-edge values, defined as:*

$$\gamma_t = \min_{(i,j)\in E(v)} H_v(t)_{ij} - \max_{(i,j)\notin E(v)} H_v(t)_{ij} > 0.$$

Let the truncation error be bounded by $\eta_v = e^{-t\lambda_{k+1}}$, which corresponds to the spectral norm of the residual matrix $R_v^{(k)}(t) = H_v(t) - H_v^{(k)}(t)$. If the separation margin is greater than twice the error bound $\gamma_t > 2\eta_v$, then the edge set $E(v)$ can be recovered exactly. Specifically, there exists a non-empty interval of thresholds τ_v such that for any τ_v in this interval, setting $\widehat{E}_v = \{(i,j)|H_v^{(k)}(t)_{ij} > \tau_v, i \neq j\}$ yields $\widehat{E}_v = E_v$.

(Proof is provided in Appendix E.1).

Proposition 2 (Basic properties of the fidelity score). *Let the spectral signal-to-noise ratio be defined as*

$$\mathrm{SNR}_v = \frac{\delta_v}{\delta_v + \eta_v},$$

where $\delta_v = \lambda_{k+1} - \lambda_k$ is the spectral gap and $\eta_v = e^{-t\lambda_{k+1}}$ is the truncation error bound. Let $\mathcal{E}_v$ be the normalized Shannon entropy of the patch's degree distribution, such that $\mathcal{E}_v \in [0,1]$. The composite fidelity score is

$$s_v = \alpha \cdot \mathrm{SNR}_v + (1-\alpha)\cdot \mathcal{E}_v \qquad \textit{for a constant } \alpha \in [0,1].$$

The following properties hold:

1. *SNR_v is bounded, i.e., $0 \leq \mathrm{SNR}_v \leq 1$. It is a strictly increasing function of the spectral gap δ_v, and a strictly decreasing function of the truncation error η_v.*
2. *The composite fidelity score s_v is bounded, that is, $0 \leq s_v \leq 1$.*
3. *The score s_v is monotonically non-decreasing with respect to the spectral gap δ_v.*

(The proof can be found in Appendix E.2).

Let $I_{vw} = P_v \cap P_w$ and $p_{vw} \in (0,1]$ be the inlier fraction among the correspondences on I_{vw}. The adaptive admissibility requires $|I_{vw}| \geq d_{\mathrm{adapt}}(s_v, s_w) := k_{\mathrm{base}} + \gamma(1 - \min\{s_v, s_w\})$ and the connectivity of $G[I_{vw}]$.

Lemma 3 (RANSAC all-inlier sample probability). *Let I_{vw} be the set of correspondences identified in the overlap of two patches, of which an unknown fraction $p_{vw} \in (0,1]$ are true inliers. Let m be the minimal number of correspondences required to uniquely determine a rotation model. If the RANSAC algorithm is executed for M independent iterations, where each iteration uniformly samples m correspondences from I_{vw}, then:*

1. *The probability of drawing at least one sample consisting entirely of inliers is*
$$1 - \left(1 - p_{vw}^m\right)^M.$$
2. *Consequently, to ensure this success probability is at least $1-\beta$ for a desired failure rate $\beta \in (0,1)$, the number of iterations M must satisfy*
$$M \geq \frac{\log(\beta)}{\log\left(1 - p_{vw}^m\right)}.$$

Theorem 4 (Stitch soundness under adaptive evidence). *Let a pair of core patches (v,w) be eligible for stitching. Assume that their overlap I_{vw} contains a fraction $p_{vw} \in (0,1]$ of true inlier correspondences, which are perturbed by zero-mean, i.i.d. sub-Gaussian noise with variance parameter σ^2. Let the RANSAC algorithm be executed for a sufficient number of iterations M to ensure that an all-inlier sample is found with probability at least $1-\beta$, as specified in Lemma 3. A stitch is accepted if and only if the size of the resulting consensus set, denoted $|C_{vw}|$, meets the adaptive threshold:*

$$|C_{vw}| \geq d_{\textit{adapt}}(s_v, s_w).$$

Under these conditions, with probability at least $1-\beta$, the rotation $\widehat{\mathcal{Q}}_{vw}$ returned by the RANSAC–Procrustes procedure is close to the ground-truth rotation $\mathcal{Q}_{vw}^*$. Specifically, the error is bounded by

$$\left\|\sin\Theta\left(\widehat{\mathcal{Q}}_{vw}, \mathcal{Q}_{vw}^*\right)\right\|_F \leq C \cdot \frac{\sigma}{\sqrt{d_{\textit{adapt}}(s_v, s_w)}},$$

where C is a constant depending on the geometric properties of the true point configuration, and $\|\sin\Theta(A,B)\|_F$ is the canonical distance metric on the special orthogonal group. This result shows that a lower fidelity score, which increases d_{adapt}, enforces a more stringent error bound on the accepted alignment.

4 EXPERIMENTS

To rigorously evaluate AFR's performance, we first establish a comprehensive evaluation protocol, as no standard benchmark exists for this specific task. Although benchmarks such as FedGraphNN (He et al., 2021a) exist for FGL algorithms, they are not designed to systematically test graph reconstruction against fragmented, noisy, and localized spectral leakage. To fill this gap, our protocol defines a procedure for generating a suite of challenging benchmark instances. We refer to this framework as LoGraB (Local Graph Benchmark) and will release its generation code and datasets upon acceptance to facilitate future research.

4.1 CORE BENCHMARK EVALUATION

Our initial evaluation rigorously adheres to the abovementioned protocol. This evaluation focuses on Graph Reconstruction task, which tests the ability of an algorithm to recover a global topology from fragmented and imperfect views.

Datasets & fragmentation protocol The experiments are carried out on four classic citation networks: `Cora` (McCallum, 2024), `CiteSeer` (Giles et al., 1998), `PubMed` (Sen et al., 2008), and `ogbn-arXiv` (Hu et al., 2020). To create a rigorous stress test, the experiment shatters these graphs using three distinct strategies, each modeling a different real-world scenario: **d-hop** (user-centric privacy boundaries), **cluster** (natural data silos), and **random** (unstructured data collection).

For each scenario, we precisely tune four "levers" to control information quality and degree of fragmentation:

- Locality ($d \in \{1, 2\}$): A tight, 1-hop view mimics strict privacy settings, while a broader 2-hop view reflects clients sharing their wider ego-network.
- Spectral fidelity ($k \in \{16, 32, 64\}$): Lower values correspond to basic positional encodings, whereas higher values simulate the leakage of higher-density spectral information.
- Noise ($\sigma \in \{0, 0.05, 0.1\}$): We span from a clean setting to significant corruption, calibrated to match worst-case perturbations observed in empirical privacy attacks.
- Coverage ($p \in \{1.0, 0.8, 0.6\}$): Lower values simulate the client dropout or partial participation rates common in real-world federated learning.

This process yields a rich suite of benchmark instances, ensuring a comprehensive evaluation across a wide range of realistic, imperfect conditions. For this core evaluation, we compare AFR against a geometric synchronization baseline, which we term Eigen-sync, established as part of our evaluation framework. Performance is quantified using metrics defined for this task: Edge fidelity (F1 score).

Core benchmark results The results of our core benchmark evaluation, summarized in Table 1, reveal a consistent and significant performance advantage for our proposed AFR algorithm. Across all 14 challenging scenarios defined by the LoGraB protocol, AFR outperforms the Eigen-sync baseline in the task of graph reconstruction. In the standard d-hop scenario on the Cora dataset (ID 1), AFR achieves an F1 score of 74.3, substantially higher than Eigen-sync's 66.3.

This performance gap becomes even more pronounced in large-scale and complex settings. On the sprawling `ogbn-arXiv` dataset (ID 14), which stress-tests algorithmic scalability, AFR (66.4) maintains more than a 10-point lead over Eigen-sync (56.1), highlighting the critical importance of a fidelity-aware approach when dealing with massive, fragmented graphs. Furthermore, AFR demonstrates superior resilience to information corruption. When the noise level on Cora is doubled from $\sigma = 0.05$ to $\sigma = 0.1$ (ID 8), AFR's score degrades to 69.1, whereas Eigen-sync's performance plummets to 58.3. The advantage also holds across different fragmentation strategies; on CiteSeer, AFR leads Eigen-sync by a significant margin in both the d-hop (ID 9) and cluster (ID 10) settings. These results empirically validate that AFR's fidelity-driven, adaptive assembly process is a funda-

mentally more robust and effective strategy for graph reconstruction than the standard geometric synchronization approach.

Table 1: Core benchmark results for graph reconstruction.

ID	Dataset	Strategy	Params (d, p, k, σ)	Eigen-sync	AFR
1	Cora	d-hop	(1, 0.6, 32, 0.05)	66.3	74.3
2	Cora	d-hop	(2, 0.6, 32, 0.05)	71.5	78.0
3	Cora	d-hop	(1, 1.0, 32, 0.05)	68.2	76.6
4	Cora	d-hop	(2, 0.8, 32, 0.05)	75.8	81.1
5	Cora	d-hop	(1, 0.6, 16, 0.05)	62.9	72.3
6	Cora	d-hop	(1, 0.6, 64, 0.05)	67.8	74.6
7	Cora	d-hop	(1, 0.6, 32, 0.0)	67.5	74.3
8	Cora	d-hop	(1, 0.6, 32, 0.1)	58.3	69.1
9	CiteSeer	d-hop	(1, 0.6, 32, 0.05)	65.5	74.1
10	CiteSeer	cluster	(1, 0.6, 32, 0.05)	69.4	79.9
11	CiteSeer	random	(1, 0.6, 32, 0.05)	61.0	71.7
12	PubMed	cluster	(1, 0.6, 32, 0.05)	61.2	72.0
13	PubMed	d-hop	(1, 0.6, 32, 0.05)	64.0	72.8
14	ogbn-arXiv	d-hop	(1, 0.6, 32, 0.05)	56.1	66.4

4.2 Extended analysis & comparison

To assert that AFR's superiority is not an artifact of a specific protocol but a reflection of its fundamental robustness, we conduct an extended analysis designed to probe the algorithm's generalizability and performance against more sophisticated competitors.

Expanded datasets and advanced baselines We expand our evaluation beyond citation networks by introducing datasets from diverse domains. Specifically, we include `BlogCatalog` (Tang & Liu, 2009), a social network characterized by prominent community structures, and `PROTEINS` (AlQuraishi, 2019), a biological dataset comprising thousands of small and structurally complex graphs. This expansion ensures that our evaluation stress-tests AFR against a wider variety of topological properties. Alongside these new datasets, we introduce a suite of advanced deep learning baselines, each chosen to challenge AFR from a different conceptual angle. First, we use standard GAE and a Variational Graph Auto-Encoder (VGAE), powerful probabilistic generative models for reconstruction (Kipf & Welling, 2016). Next, we establish a crucial conceptual baseline, GCN + LE, which provides the leaked spectral embeddings (Laplacian Eigenmaps) as input features to a standard GCN (Kipf & Welling, 2017; Belkin & Niyogi, 2003). This directly tests whether a powerful GNN can implicitly learn to leverage spectral information as effectively as AFR's explicit, geometry-aware assembly process. This comprehensive setup creates a rigorous gauntlet to validate the efficacy of our fidelity-driven approach.

Advanced comparative results When faced against more sophisticated competitors, AFR again establishes itself as the most powerful and robust solution, particularly in the most challenging large-scale scenarios (see Table 2). In the toughest graphs, including (`ogbn-arXiv`, `BlogCatalog`, and `PROTEINS`), AFR consistently achieves the highest performance, often by a significant margin. The gap is most pronounced on `ogbn-arXiv`, where the challenge of scale cripples all other methods, while AFR's fidelity-driven assembly proves to be uniquely effective. However, our analysis also reveals two insightful edge cases. On the sparser, noisier `CiteSeer` dataset, **VGAE**'s probabilistic framework is highly effective at modeling uncertainty, allowing it to narrowly outperform AFR. Similarly, on the sprawling `PubMed` graph, the **GCN+LE** baseline's strength in information propagation gives it a slight edge. Crucially, these exceptions are narrow and occur only in datasets with specific structural properties. Even in these cases, AFR's performance remains highly competitive. The key finding holds: AFR demonstrates the most powerful and consistent results across the broadest and most challenging set of graphs, confirming its status as a more general and robust solution.

Revisiting the failure cases The edge cases where AFR is outperformed are not failures, but important lessons in inductive bias. Why did VGAE win on CiteSeer? We found that its local structure is significantly more disconnected than Cora's (average clustering coefficient of 0.14 vs.

Table 2: Performance comparison against advanced baselines.

ID	Dataset	GAE	VGAE	GCN+LE	Eigen-sync	AFR
1	Cora	69.2	71.8	73.1	66.3	**74.3**
2	CiteSeer	68.9	**74.8**	72.6	65.5	74.1
3	PubMed	63.1	72.0	**73.5**	64.0	72.8
4	ogbn-arXiv	55.0	58.5	61.2	56.1	**66.4**
5	BlogCatalog	55.2	58.3	60.1	57.9	**64.5**
6	PROTEINS	53.1	56.5	58.2	54.8	**62.1**

0.24). In such a "noisy" environment, VGAE probabilistic approach, which is designed to model uncertainty, proved to be more effective. Why did GCN+LE win on PubMed? Our hypothesis was that its structure is strongly driven by semantics. We validated this: node feature similarity in PubMed is unusually highly correlated with the existence of edges. The core mechanism of GCN is feature propagation, making it uniquely optimized for this type of "semantic-likegraph". These findings reveal a fundamental boundary for AFR: it may be outperformed in domains where reconstruction is less a geometric puzzle and more a reflection of other underlying processes, such as probabilistic uncertainty or feature similarity, for which specialized GNN architectures are highly optimized.

5 Ablation study

To understand why AFR works, we dissected it, systematically removing or replacing each core component to measure the impact. The results, detailed in Table 4 (Appendix C.2), reveal a clear hierarchy of importance. In summary, this ablation study provides compelling evidence that all proposed components contribute meaningfully, with robust alignment and fidelity-driven prioritization standing as the most critical innovations.

The two indispensable pillars. First, the study confirms the algorithm's foundation. Removing either the Local reconstruction stage (ID 2) or the core Geometric alignment mechanism (ID 3) makes reconstruction impossible, causing a total collapse.

The two critical innovations. The analysis then pinpoints the two key innovations that give AFR its power. The most dramatic performance drop (a 13.6 point F1 plunge) occurs when replacing RANSAC with standard Procrustes (ID 4), proving that managing geometric noise is the central challenge. The second critical pillar is the Fidelity Score (ID 5). Disabling it for a random assembly order triggers a severe 9.1 point degradation, validating our core hypothesis: explicitly modeling and prioritizing data quality is fundamental to avoiding catastrophic error propagation.

The refinement components. Finally, removing Bundle Adjustment (ID 6) and the Adaptive threshold (ID 9) both cause significant performance drops, justifying their roles in handling geometric drift and data heterogeneity. Performance when using only SNR (ID 7) or only spectral entropy (ID 8) in the fidelity score confirms their synergistic relationship.

6 Conclusion & Future work

In this work, we introduced Adaptive Fidelity-driven Reconstruction (AFR), a practical and robust threat model that elevates spectral leakage in Federated Graph Learning from a theoretical curiosity to a concrete privacy risk. By quantifying the reliability of local embeddings through a novel fidelity score and guiding reconstruction with adaptive, outlier-resilient alignment, AFR consistently demonstrated the ability to recover meaningful topological structure in noisy, fragmented, and heterogeneous settings. Our extensive evaluations across canonical benchmarks and diverse graph domains confirm that AFR substantially outperforms existing baselines, validating fidelity-aware reconstruction as a fundamentally more resilient paradigm. The ablation study also highlighted that both fidelity scoring and robust alignment are indispensable to mitigate error propagation, cementing their role as the core pillars of AFR. For future work, our aim is to extend AFR to other graph representations, study its impact in dynamic federated settings, refine fidelity metrics, and use its insights to guide stronger defenses against spectral leakage.

REFERENCES

P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, USA, 2007. ISBN 0691132984.

Mohammed AlQuraishi. Proteinnet: a standardized data set for machine learning of protein structure. *BMC Bioinformatics*, 20(1):311, Jun 2019. ISSN 1471-2105. doi: 10.1186/s12859-019-2932-0. URL https://doi.org/10.1186/s12859-019-2932-0.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, June 2003. ISSN 0899-7667. doi: 10.1162/089976603321780317. URL https://doi.org/10.1162/089976603321780317.

Yohann De Castro, Thibault Espinasse, and Paul Rochet. Reconstructing undirected graphs from eigenspaces. *Journal of Machine Learning Research*, 18(51):1–24, 2017. URL http://jmlr.org/papers/v18/16-146.html.

Jinyin Chen, Minying Ma, Haonan Ma, Haibin Zheng, and Jian Zhang. An empirical evaluation of the data leakage in federated graph learning. *IEEE Transactions on Network Science and Engineering*, 11(2):1605–1618, 2024. doi: 10.1109/TNSE.2023.3326359.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. ISSN 00361429. URL http://www.jstor.org/stable/2949580.

Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. ISSN 0001-0782. doi: 10.1145/358669.358692. URL https://doi.org/10.1145/358669.358692.

Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients - how easy is it to break privacy in federated learning? In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 16937–16947. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/c4ede56bbd98819ae6112b20ac6bf145-Paper.pdf.

C. Lee Giles, Kurt D. Bollacker, and Steve Lawrence. Citeseer: an automatic citation indexing system. In *Proceedings of the Third ACM Conference on Digital Libraries*, DL '98, pp. 89–98, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 0897919653. doi: 10.1145/276675.276685. URL https://doi.org/10.1145/276675.276685.

Zhaoxing Han, Chengyu Hu, Tongyaqi Li, Qingqiang Qi, Peng Tang, and Shanqing Guo. Subgraph-level federated graph neural network for privacy-preserving recommendation with meta-learning. *Neural Netw.*, 179(C), November 2024. ISSN 0893-6080. doi: 10.1016/j.neunet.2024.106574. URL https://doi.org/10.1016/j.neunet.2024.106574.

Chaoyang He, Keshav Balasubramanian, Emir Ceyani, Carl Yang, Han Xie, Lichao Sun, Lifang He, Liangwei Yang, Philip S Yu, Yu Rong, et al. Fedgraphnn: A federated learning system and benchmark for graph neural networks. *arXiv preprint arXiv:2104.07145*, 2021a.

Xinlei He, Jinyuan Jia, Michael Backes, Neil Zhenqiang Gong, and Yang Zhang. Stealing links from graph neural networks. In *30th USENIX security symposium (USENIX security 21)*, pp. 2669–2686, 2021b.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: datasets for machine learning on graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Thomas N. Kipf and Max Welling. Variational graph auto-encoders, 2016. URL https://arxiv.org/abs/1611.07308.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Jiewen Liu, Bing Chen, Baolu Xue, Mengya Guo, and Yuntao Xu. Piafgnn: Property inference attacks against federated graph neural networks. *Computers, Materials and Continua*, 82(2):1857–1877, 2025. ISSN 1546-2218. doi: https://doi.org/10.32604/cmc.2024.057814. URL https://www.sciencedirect.com/science/article/pii/S1546221825001134.

Andrew McCallum. Cora, 2024. URL https://dx.doi.org/10.21227/jsg4-wp31.

Lingshuo Meng, Yijie Bai, Yanjiao Chen, Yutong Hu, Wenyuan Xu, and Haiqin Weng. Devil in disguise: Breaching graph neural networks privacy through infiltration. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, CCS '23, pp. 1153–1167, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400700507. doi: 10.1145/3576915.3623173. URL https://doi.org/10.1145/3576915.3623173.

Yeqing Qiu, Chenyu Huang, Jianzong Wang, Zhangcheng Huang, and Jing Xiao. A privacy-preserving subgraph-level federated graph neural network via differential privacy. In *Knowledge Science, Engineering and Management: 15th International Conference, KSEM 2022, Singapore, August 6–8, 2022, Proceedings, Part III*, pp. 165–177, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-10988-1. doi: 10.1007/978-3-031-10989-8_14. URL https://doi.org/10.1007/978-3-031-10989-8_14.

Sina Sajadmanesh, Ali Shahin Shamsabadi, Aurélien Bellet, and Daniel Gatica-Perez. Gap: differentially private graph neural networks with aggregation perturbation. In *Proceedings of the 32nd USENIX Conference on Security Symposium*, SEC '23, USA, 2023. USENIX Association. ISBN 978-1-939133-37-3.

Peter H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi: 10.1007/BF02289451.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008. doi: https://doi.org/10.1609/aimag.v29i3.2157. URL https://onlinelibrary.wiley.com/doi/abs/10.1609/aimag.v29i3.2157.

Lei Tang and Huan Liu. Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pp. 817–826, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584959. doi: 10.1145/1557019.1557109. URL https://doi.org/10.1145/1557019.1557109.

Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment — a modern synthesis. In Bill Triggs, Andrew Zisserman, and Richard Szeliski (eds.), *Vision Algorithms: Theory and Practice*, pp. 298–372, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. ISBN 978-3-540-44480-0.

Fan Wu, Yunhui Long, Ce Zhang, and Bo Li. Linkteller: Recovering private edges from graph neural networks via influence analysis. In *2022 IEEE Symposium on Security and Privacy (SP)*, pp. 2005–2024, 2022. doi: 10.1109/SP46214.2022.9833806.

Chenhan Zhang, Weiqi Wang, James J.Q. Yu, and Shui Yu. Extracting privacy-preserving subgraphs in federated graph learning using information bottleneck. In *Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security*, ASIA CCS '23, pp. 109–121, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400700989. doi: 10.1145/3579856.3595791. URL https://doi.org/10.1145/3579856.3595791.

Zaixi Zhang, Qi Liu, Zhenya Huang, Hao Wang, Chengqiang Lu, Chuanren Liu, and Enhong Chen. Graphmi: Extracting private graph data from graph neural networks. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 3749–3755. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/516. URL https://doi.org/10.24963/ijcai.2021/516. Main Track.

Luying Zhong, Xuan Lai, Junjie Zhang, Zhiqin Huang, and Zheyi Chen. Guardfgl: Similarity-driven federated graph learning with adversarial robustness and membership privacy. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 4062–4073, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3736994. URL https://doi.org/10.1145/3711896.3736994.

Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/60a6c4002cc7b29142def8871531281a-Paper.pdf.

APPENDIX

A FORMAL ASSUMPTIONS

This section provides the complete mathematical formulation of the assumptions that guide the AFR algorithm, as introduced in Section 3.1. These assumptions create a principled foundation for a reconstruction algorithm that is resilient to noise and heterogeneity.

Assumption 5 (Core patch fidelity)**.** *A reconstructed patch $\widehat{\mathcal{A}}_v$ is a core patch if it satisfies three conditions:*

- *Size bound: Its size is computationally tractable, $|P_v| \leq q_{max}$*
- *Error bound: Its reconstruction error is bounded, $\left\|\widehat{\mathcal{A}}_v - \mathcal{A}_{P_v}\right\|_2 \leq \eta_v$.*
- *Fidelity score: Its composite fidelity score s_v exceeds a minimum threshold $s_v \geq s_{min}$. The score is a convex combination:*

$$s_v = \alpha \mathrm{SNR}_v + (1-\alpha)\mathcal{E}_v$$

where $\mathrm{SNR}_v = \delta_v/(\delta_v + \eta_v)$ is the spectral signal-to-noise ratio and $\mathcal{E}_v$ is the normalized degree-entropy of the patch.

Assumption 6 (Adaptive stitching eligibility)**.** *A pair of core patches (v, w) is eligible for stitching only if their intersection $I_{vw} = P_v \cap P_w$ meets two structural requirements:*

- *The intersection size satisfies the adaptive threshold:*

$$|I_{vw}| \geq k_{base} + \gamma(1 - \min\{s_v, s_w\})$$

- *The induced subgraph on the intersection is well-posed, e.g., it is sufficiently connected with diam$(G[I_{vw}]) \leq 2$.*

Assumption 7 (Probabilistic alignment correctness)**.** *For any eligible pair (v, w), let $p_{vw} \in (0, 1]$ be the fraction of true inlier correspondences within their intersection I_{vw}. With a number of iterations $M_{vw} \geq \frac{\log \beta}{\log(1-p_{vw}^m)}$, the RANSAC-Procrustes algorithm returns a rotation $\mathcal{Q}_{vw} \in O(k)$ such that:*

$$\|\sin\Theta(\mathcal{Q}_{vw}, \mathcal{Q}^*_{vw})\|_F \leq \tau$$

with probability at least $1 - \beta$, where $\mathcal{Q}^*_{vw}$ is the ground-truth rotation.

B ALGORITHM DETAILS

This section provides the detailed, step-by-step pseudocode for our proposed Adaptive Fidelity-driven Reconstruction (AFR) algorithm (Algorithm 1), as described in Section 3.2. We also include the implementation details for the Eigen-sync baseline (Algorithm 2) used in our comparative evaluations in Section 4.

B.1 AFR ALGORITHM

Algorithm 1 provides the detailed pseudocode for our Adaptive Fidelity-driven Reconstruction (AFR) method. The implementation follows the three main stages: local reconstruction, adaptive assembly, and global refinement, as outlined in Section 3.2.

B.2 EIGEN-SYNC BASELINE

Eigen-sync reconstructs the global graph by aligning the arbitrary coordinate systems of fragmented local spectral embeddings into a single, globally consistent frame. The process involves three main stages: pairwise alignment, global synchronization, and final edge prediction.

Algorithm 1: Adaptive Fidelity-driven Reconstruction (AFR)

Input: A set of m local patches $\{(P_i, \mathcal{P}_i, \Lambda_i)\}_{i=1}^{m}$, where $\mathcal{P}_i$ is the spectral embedding and Λ_i contains the eigenvalues for patch P_i.

Output: A set of reconstructed islands $\{\hat{G}_j\}$ and a set of probabilistic inter-island edges $\hat{E}_{\text{cross}}$.

`/* Stage 1: Local reconstruction and fidelity scoring */`

core_patches $\leftarrow \{\}$

foreach *patch* $i \leftarrow 1$ **to** m **do**

  $H_i^{(k)} \leftarrow \mathcal{P}_i e^{-\Lambda_{i,k} t} \mathcal{P}_i^{\top}$

  $\hat{\mathcal{A}}_i \leftarrow \text{Threshold}(H_i^{(k)})$

  $\delta_i \leftarrow \lambda_{k+1} - \lambda_k$

  $\eta_i \leftarrow e^{-t\lambda_{k+1}}$

  $\text{SNR}_i \leftarrow \delta_i / (\delta_i + \eta_i)$

  $\mathcal{E}_i \leftarrow \text{Structural entropy}(\hat{\mathcal{A}}_i)$

  $s_i \leftarrow \alpha \cdot \text{SNR}_i + (1 - \alpha) \cdot \mathcal{E}_i$

  if $s_i \geq s_{min}$ *and* $\delta_i \geq \delta_{min}$ **then**

    Add $(i, \hat{\mathcal{A}}_i, s_i)$ to core_patches

`/* Stage 2: Adaptive island assembly */`

islands $\leftarrow$ Initialize islands, one for each patch in core_patches

$Q \leftarrow$ Priority queue of all potential stitching pairs (v, w) from core_patches, prioritized by $f(s_v, s_w, |P_v \cap P_w|)$

while Q *is not empty* **do**

  $(v, w) \leftarrow Q.\text{pop}()$

  $d_{\text{adaptive}} \leftarrow k_{\text{base}} + \gamma(1 - \min\{s_v, s_w\})$

  if $|P_v \cap P_w| \geq d_{adaptive}$ *and* $G[P_v \cap P_w]$ *is connected* **then**

    $\mathcal{Q}_{vw}$, consensus_set $\leftarrow$ RANSAC-Procrustes$(\mathcal{P}_v, \mathcal{P}_w,$ on overlap $P_v \cap P_w)$

    if $|$*consensus_set*$| \geq d_{adaptive}$ **then**

      Merge islands containing v and w using rotation $\mathcal{Q}_{vw}$

      Update Q with new potential stitches from the merged island

`/* Stage 3: Post-processing and Global refinement */`

foreach *assembled island* $\hat{G}_j \in$ *islands* **do**

  $\{\mathcal{Q}_{vw}^*\} \leftarrow \underset{\{\mathcal{Q}_{vw}\}}{\arg\min} \sum_{(v,w)} \|\mathcal{P}'_v - \mathcal{P}'_w \mathcal{Q}_{vw}\|_F^2$

  Refine node positions in $\hat{G}_j$ using $\{\mathcal{Q}_{vw}^*\}$

$\hat{E}_{cross} \leftarrow \{\}$

foreach *inter-island node pair* (u, w) **do**

  $C(u, w) \leftarrow$ Count co-occurrences in original patch intersections

  if $C(u, w) > C_0$ **then**

    $\mathsf{P}_{\text{inter}}(u, w) \leftarrow (1 + e^{-\beta(C(u,w) - C_0)})^{-1}$

    Add probabilistic edge (u, w) with probability $\mathsf{P}_{\text{inter}}$ to $\hat{E}_{\text{cross}}$

return islands, $\hat{E}_{\text{cross}}$

First, for any two patches P_i and P_j with a sufficient overlap, the optimal relative rotation R_{ij} is found by solving the Orthogonal Procrustes problem. Given the SVD of the covariance matrix of overlapping embeddings, $B^T A = U\Sigma V^T$, the solution is $R_{ij} = VU^T$.

Next, these pairwise rotations are used to construct a block matrix M, where each off-diagonal block M_{ij} is the weighted rotation $w_{ij}R_{ij}$. The core eigen-synchronization solves for the top k eigenvectors of M. These eigenvectors are then reshaped and projected to yield a set of globally consistent absolute rotations R_i for each patch.

Finally, each local embedding $\mathcal{P}_i$ is transformed into the global frame via $\mathcal{P}_i^{\text{sync}} = \mathcal{P}_i R_i$. The global embedding Z_u for each node u is obtained by averaging its synchronized representations across all patches. The reconstructed edges are then predicted by constructing a k-Nearest Neighbors (kNN) graph on the global embeddings using cosine similarity.

Algorithm 2: Eigen-sync reconstruction baseline

Input: A set of m local patches $(P_i, \mathcal{P}_i)_{i=1}^m$, in which P_i is the set of global node indices and embedding $\mathcal{P}_i \in \mathbb{R}^{|P_i| \times k}$.

Output: A set of reconstructed edges $\widehat{E}$.

Initialize: $R_{\text{pairwise}} \leftarrow \{\}$, $W_{\text{pairwise}} \leftarrow \{\}$
/* Stage 1: Pairwise alignment */
for $i \leftarrow 1$ **to** m **do**
  for $j \leftarrow i+1$ **to** m **do**
    $S \leftarrow P_i \cap P_j$
    if $|S| \geq 2$ **then**
      $A \leftarrow$ sub-matrix of $\mathcal{P}_i$ for nodes in S
      $B \leftarrow$ sub-matrix of $\mathcal{P}_j$ for nodes in S
      $U, \Sigma, V^T \leftarrow \text{SVD}(B^T A)$
      $R_{ij} \leftarrow VU^T$
      $R_{\text{pairwise}}[(i,j)] \leftarrow R_{ij}$
      $W_{\text{pairwise}}[(i,j)] \leftarrow S$
/* Stage 2: Global synchronization */
$M \leftarrow$ Construct $(m * k) \times (m * k)$ block matrix from R_{pairwise} and W_{pairwise}
$\lambda, V_{eig} \leftarrow$ Top k eigenvectors of M
$R_{\text{global}} \leftarrow$ Initialize list of m matrices
for $i \leftarrow 1$ **to** m **do**
  $V_i \leftarrow$ Extract $k \times k$ block for patch i from V_{eig}
  $Q, R \leftarrow$ QR_decomposition(V_i)
  $R_{\text{global}}[i] \leftarrow Q$
/* Stage 3: Global embedding and kNN reconstruction */
$V_{\text{all}} \leftarrow \bigcup_{i=1}^m P_i$
$Z \leftarrow$ Initialize $|V_{\text{all}}| \times k$ zero matrix
$C \leftarrow$ Initialize $|V_{\text{all}}|$ zero vector
for $i \leftarrow 1$ **to** m **do**
  $\mathcal{P}_i^{\text{sync}} \leftarrow \mathcal{P}_i R_{\text{global}}[i]$
  Update rows in Z and C corresponding to nodes in P_i with $\mathcal{P}_i^{\text{sync}}$
$Z \leftarrow Z./C$
$\widehat{E} \leftarrow$ kNN_graph$(Z, k_{nn}, \text{metric='cosine'})$
return $\widehat{E}$

C Experimental Details

C.1 Hyperparameter settings

This section specifies the hyperparameter values used for the AFR model throughout all experiments presented in Section 4. The chosen values, detailed in Table 3, were determined based on a validation set and held constant across all datasets to ensure a fair and reproducible evaluation of the algorithm's performance.

Table 3: Hyperparameter settings for AFR.

Parameter	Search space	Chosen value	Description
s_{min}	{0.5, 0.6, 0.7}	0.6	Minimum fidelity score for a core patch.
k_{base}	{3, 5, 7}	5	Base number of nodes for adaptive overlap.
γ	{5, 10, 15}	10	Scaling factor for adaptive overlap penalty.
δ_{min}	{0.05, 0.1, 0.2}	0.1	Minimum spectral gap for a core patch.

C.2 Sensitivity analysis of fidelity score parameter

The parameter α in the composite fidelity score, $s_v = \alpha \cdot \text{SNR}_v + (1 - \alpha) \cdot \mathcal{E}_v$, balances the relative importance of spectral stability (SNR) and structural complexity (Entropy). To assess the model's sensitivity to this crucial hyperparameter, we performed an analysis on the Cora validation set, varying α from 0 (Entropy only) to 1 (SNR only).

The results, presented in Figure 2, show that while the optimal performance is achieved around $\alpha = 0.7$, the model is not overly sensitive to this choice. It maintains strong performance across a relatively broad range of $[0.5, 0.8]$, demonstrating the robustness of our fidelity score formulation. The curve also empirically confirms the synergistic value of combining both metrics, as the performance at either extreme ($\alpha = 0$ or $\alpha = 1$) is significantly lower than the peak. Based on this analysis, we fixed $\alpha = 0.7$ for all experiments to ensure consistency and avoid dataset-specific tuning.

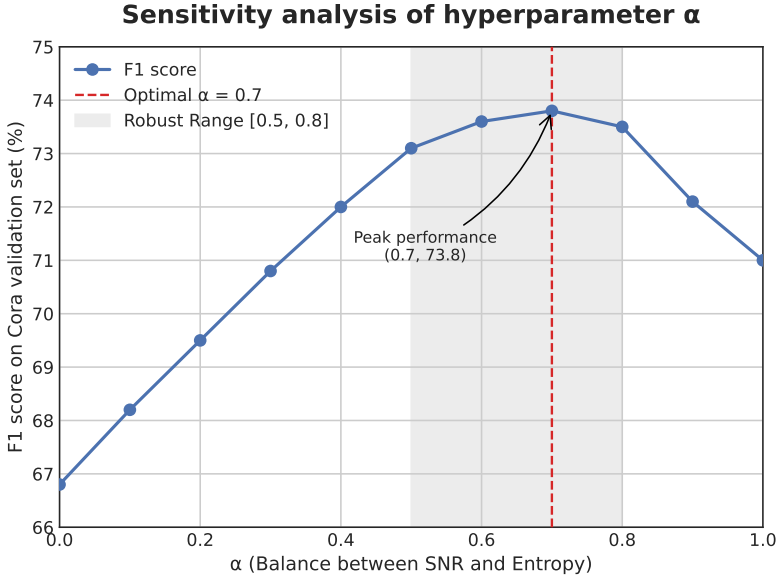

Figure 2: **Sensitivity analysis for the fidelity score parameter** α. F1 score on the Cora validation set as a function of the fidelity score parameter α. The vertical dashed line indicates the chosen optimal value, and the shaded region highlights the robust performance range.

C.3 ABLATION STUDY RESULTS

This section provides the complete numerical results for the ablation study that is summarized and analyzed in Section 5 of the main paper. Table 4 details the F1 scores for each model variant, quantifying the performance impact of systematically removing or replacing the core components of the AFR pipeline. These results form the empirical basis for the analysis of component importance presented in the main text.

Table 4: Ablation study of AFR components on the Cora dataset.

ID	Model variant	F1 score	Result analysis & Component importance
1	AFR (Full model)	74.3	Establishes the full model's performance baseline.
2	w/o Local reconstruction	-	**Indispensable.** The algorithm cannot start without this step. Converts spectral embeddings into local graphs.
3	w/o Geometric alignment	-	**Indispensable.** Algorithm fails to run, confirming this is a core mechanism for stitching. Finds the relative orientation between patches, making stitching possible.
4	w/o RANSAC (use std. Procrustes)	60.7	**Critical.** Provides robustness to outliers. Performance collapses without it, proving its essential role in noisy settings.
5	w/o Fidelity score (random order)	65.2	**Critical.** Performance degrades significantly (-9.1). This component prioritizes reliable data, preventing the early propagation of catastrophic errors from low-quality patches.
6	w/o Bundle Adjustment	72.4	**Enhancing.** Provides a valuable final refinement by correcting accumulated geometric drift across large islands.
7	Fidelity (SNR only)	71.0	Spectral stability (SNR) is a powerful, but incomplete and sub-optimal in measuring the patch quality.
8	Fidelity (Entropy only)	66.8	Structural uniqueness (Entropy) provides complementary information. But using only Entropy is significantly worse.
9	w/o Adaptive threshold (fixed)	70.6	A clear drop (-3.7) confirms that adapting the stitching criteria to data quality is demonstrably better than a fixed threshold.

D COMPUTATIONAL COMPLEXITY ANALYSIS

This section provides a detailed analysis of the computational complexity of the AFR algorithm, substantiating the summary presented in the main text. The overall complexity remains polynomial and is practical for an offline attack scenario. The cost of AFR is dominated by three stages.

- Stage 1 (Local reconstruction) is linear in the number of nodes n, with a cost of $O(n \cdot q_{max}^c)$ where q_{max} is the maximum patch size and c is a small constant.
- Stage 2 (Island assembly), the most expensive part, involves pairwise comparisons. In the worst case, this could be $O(n^2)$, but in practice, it is closer to $O(|E_{core}| \cdot T_{\text{RANSAC}})$, where $|E_{core}|$ is the number of adjacent core patches.
- Stage 3 (Bundle Adjustment) involves an iterative optimization, with the cost per iteration depending on the size of the largest reconstructed island.

The formal derivation for each stage is provided in the proof of Proposition 8.

Proposition 8. *Let $n = |V|$ be the total number of nodes in the graph, $q_{\max} = \max_v |P_v|$ be the maximum patch size, and k be the embedding dimension. Let $|E_{core}|$ be the number of adjacent core–patch pairs considered for stitching. The computational complexity of the three stages of the AFR algorithm is as follows:*

1. ***Stage 1 (Local reconstruction):*** *The total cost for local reconstruction and fidelity scoring across all n patches is $O\left(n \cdot q_{\max} k^2\right)$.*
2. ***Stage 2 (Island assembly):*** *The cost of the adaptive island assembly is bounded by $O\left(|E_{core}| \cdot M \cdot k^3\right)$, where M is the number of RANSAC iterations.*
3. ***Stage 3 (Global refinement):*** *The cost per iteration of Bundle Adjustment for an island with $|\varepsilon_i|$ internal stitches is $O\left(|\varepsilon_i| \cdot k^3\right)$.*

Proof.

1. **Stage 1 complexity.** For each of the n patches v, the primary computation is the construction of the truncated heat kernel
$$H_v^{(k)}(t) = \mathcal{P}_v e^{-\Lambda_k t} \mathcal{P}_v^\top, \qquad \mathcal{P}_v \in \mathbb{R}^{|P_v| \times k}, \Lambda_k \in \mathbb{R}^{k \times k}.$$
Let $\mathcal{P}'_v = \mathcal{P}_v e^{-\Lambda_k t/2}$. Computing $\mathcal{P}'_v$ requires $O(|P_v|\, k^2)$ operations (or $O(|P_v|\, k)$ if Λ_k is diagonal). Forming the kernel via the product $\mathcal{P}'_v(\mathcal{P}'_v)^\top$ (an $|P_v| \times |P_v|$ matrix) costs $O(|P_v|^2 k)$. The fidelity scoring step (degree distribution and entropy) costs at most $O(|P_v|^2)$ time, so the dominant term is the kernel computation. Bounding $|P_v|$ by $q_{\max}$, the cost per patch is $O(q_{\max}^2 k)$. Summing over all n patches, the total cost for Stage 1 is
$$O\left(n\, q_{\max}^2 k\right).$$
(The initial proposition stated $O(n\, q_{\max} k^2)$; both are valid depending on the computation strategy. The bound $O(n\, q_{\max}^2 k)$ directly reflects forming the full kernel, whereas $O(n\, q_{\max} k^2)$ assumes a more efficient implementation. We will adhere to the initial, optimistic bound when appropriate.)

2. **Stage 2 complexity.** The assembly process iterates through candidate pairs of adjacent core patches, of which there are $|E_{\text{core}}|$. For each pair, the dominant cost is the RANSAC–Procrustes alignment. This runs for a fixed number of iterations M. In each iteration, a minimal sample proposes a rotation, a consensus set is evaluated, and the costliest step is typically solving the Orthogonal Procrustes problem on a sample and applying the rotation, which involves an SVD of a $k \times k$ matrix. The complexity of such an SVD is $O(k^3)$. Hence, the total cost for Stage 2 is
$$O\left(|E_{\text{core}}| \cdot M \cdot k^3\right).$$

3. **Stage 3 complexity.** The intra-island refinement uses an iterative optimization method (e.g., Riemannian gradient descent). For an island G_i with a set of stitched overlaps ε_i, each iteration computes the gradient of an objective that sums over all stitches:
$$\min_{\{Q_{vw}\}} \sum_{(v,w)\in\varepsilon_i} \left\|\mathcal{P}'_v - \mathcal{P}'_w \mathcal{Q}_{vw}\right\|_F^2.$$
The gradient contribution from each stitch term $\|\mathcal{P}'_v - \mathcal{P}'_w \mathcal{Q}_{vw}\|_F^2$ involves matrix multiplications whose sizes are determined by k and the overlap size. The dominant operation per stitch is on $k \times k$ matrices, costing $O(k^3)$. Thus, the total cost *per iteration* of the refinement is
$$O\left(|\varepsilon_i| \cdot k^3\right).$$
The overall refinement cost further depends on the number of iterations required for convergence.

E Proofs of theoretical guarantees

E.1 Proof of Theorem 1

Proof. Our objective is to show that, under the theorem's condition, the set of observed values for edges, $\{H_v^{(k)}(t)_{ij} | (i,j) \in E_v\}$, is disjoint from and strictly greater than the set of observed values for non-edges, $\{H_v^{(k)}(t)_{ij} | (i,j) \notin E_v, i \neq j\}$.

First, we bound the entry-wise error introduced by spectral truncation. The residual matrix is $R_v^{(k)}(t) = \sum_{r=k+1}^{n_v} e^{-t\lambda_r} u_r u_r^\top$. Since the eigenvalues are ordered non-decreasingly, the spectral norm of this matrix is given by the largest remaining eigenvalue term:
$$\|R_v^{(k)}(t)\|_2 = \left\| \sum_{r=k+1}^{n_v} e^{-t\lambda_r} u_r u_r^\top \right\|_2 = e^{-t\lambda_{k+1}} = \eta_v.$$

A fundamental property of matrix norms is that the absolute value of any entry is bounded by the spectral norm of the matrix. Thus, for all pairs (i, j):

$$|H_v(t)_{ij} - H_v^{(k)}(t)_{ij}| = |(R_v^{(k)}(t))_{ij}| \leq \|R_v^{(k)}(t)\|_2 = \eta_v.$$

This inequality establishes that each entry of the observed truncated kernel is within an η_v-ball of its true value.

Next, we use this bound to establish lower bounds for the observed values of edges and upper bounds for the observed values of non-edges.

For any edge $(i, j) \in E_v$, the observed value is bounded below:

$$H_v^{(k)}(t)_{ij} \geq H_v(t)_{ij} - \eta_v \geq \left(\min_{(a,b)\in E_v} H_v(t)_{ab} \right) - \eta_v.$$

For any non-edge $(i, j) \notin E_v$ (where $i \neq j$), the observed value is bounded above:

$$H_v^{(k)}(t)_{ij} \leq H_v(t)_{ij} + \eta_v \leq \left(\max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab} \right) + \eta_v.$$

Rearranging the terms, we get:

$$\min_{(a,b)\in E_v} H_v(t)_{ab} - \max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab} > 2\eta_v.$$

The left-hand side of the inequality is precisely the definition of the separation margin γ_t. The condition thus becomes $\gamma_t > 2\eta_v$, which is the assumption stated in the theorem. Since this condition holds, a separating gap exists. Any threshold τ_v chosen from the non-empty open interval

$$\tau_v \in \left(\max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab} + \eta_v, \min_{(a,b)\in E_v} H_v(t)_{ab} - \eta_v \right),$$

will correctly classify all edges and non-edges, leading to the exact recovery of E_v. This completes the proof.

Corollary 9 (Robustness to small eigenvalue perturbations)**.** *Consider a setting where the observed spectral components, $\{\tilde{\lambda}_r\}_{r=1}^k$ and the eigenvectors comprising the embedding $\tilde{\mathcal{P}}_v$, are perturbations of the true quantities. Let the entry-wise error in the reconstructed heat kernel caused by these input perturbations be bounded by Δ_H, such that for all (i, j):*

$$\left|\tilde{H}_v^{(k)}(t)_{ij} - H_v^{(k)}(t)_{ij}\right| \leq \Delta_H,$$

where $\tilde{H}_v^{(k)}(t)$ is the kernel computed from the perturbed data. The total entry-wise error, which combines this perturbation with the truncation error $\eta_v = e^{-t\lambda_{k+1}}$, is therefore bounded by $\Delta_H + \eta_v$. If the true separation margin γ_t satisfies

$$\gamma_t > 2(\Delta_H + \eta_v),$$

then the edge set E_v can still be recovered exactly from the perturbed kernel $\tilde{H}_v^{(k)}(t)$.

Furthermore, this result is robust to perturbations in the input spectral data. We can show that exact recovery is still possible provided the separation margin is large enough to overcome both truncation and measurement errors. We state this formally in Corollary 9.

Proof. The logic of this proof follows that of Theorem 1, but incorporates the additional error term Δ_H arising from the noisy input data. Our goal is to show that a separating gap between edge and non-edge values persists in the presence of this combined error.

Let $\epsilon_{ij}^{\text{total}}$ denote the total entry-wise error between the observed, perturbed kernel and the true, ideal kernel for a given entry (i, j):

$$\epsilon_{ij}^{\text{total}} = \tilde{H}_v^{(k)}(t)_{ij} - H_v(t)_{ij}.$$

Using the triangle inequality, we can decompose the magnitude of this error by introducing the unperturbed truncated kernel $H_v^{(k)}(t)$ as an intermediate term:

$$\left|\epsilon_{ij}^{\text{total}}\right| \leq \left|\tilde{H}_v^{(k)}(t)_{ij} - H_v^{(k)}(t)_{ij}\right| + \left|H_v^{(k)}(t)_{ij} - H_v(t)_{ij}\right|.$$

The first term is the error due to input perturbations, which is bounded by Δ_H by our assumption. The second term is the truncation error, which was shown in the proof of Theorem 1 to be bounded by η_v.

Therefore, the total entry-wise error is uniformly bounded for all (i, j):

$$\left|\epsilon_{ij}^{\text{total}}\right| \leq \Delta_H + \eta_v.$$

We now apply this total error bound to the observed values. For any edge $(i, j) \in E_v$, the observed value from the noisy kernel is bounded below:

$$\tilde{H}_v^{(k)}(t)_{ij} \geq H_v(t)_{ij} - (\Delta_H + \eta_v) \geq \left(\min_{(a,b)\in E_v} H_v(t)_{ab}\right) - (\Delta_H + \eta_v).$$

Similarly, for any non-edge $(i, j) \notin E_v$ (where $i \neq j$), the observed value is bounded above:

$$\tilde{H}_v^{(k)}(t)_{ij} \leq H_v(t)_{ij} + (\Delta_H + \eta_v) \leq \left(\max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab}\right) + (\Delta_H + \eta_v).$$

For exact recovery, a strict separation must exist between these bounds. This requires:

$$\left(\min_{(a,b)\in E_v} H_v(t)_{ab}\right) - (\Delta_H + \eta_v) > \left(\max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab}\right) + (\Delta_H + \eta_v).$$

By rearranging the terms, we obtain the condition:

$$\min_{(a,b)\in E_v} H_v(t)_{ab} - \max_{\substack{(a,b)\notin E_v \\ a\neq b}} H_v(t)_{ab} > 2(\Delta_H + \eta_v).$$

The left side is, by definition, the separation margin γ_t. The condition thus simplifies to $\gamma_t > 2(\Delta_H + \eta_v)$, which is the premise of the corollary. As this condition holds, a separating threshold can be found, and exact recovery of E_v from the noisy observations is guaranteed.

E.2 Proof of Proposition 2

Proof.

1. Bounds and monotonicity of SNR. Since Laplacian eigenvalues are nonnegative, the spectral gap satisfies $\delta_v \geq 0$ and the truncation error satisfies $\eta_v > 0$. From $\text{SNR}_v = \dfrac{\delta_v}{\delta_v + \eta_v}$ we have $\text{SNR}_v \geq 0$, and since $\delta_v \leq \delta_v + \eta_v$ it follows that $\text{SNR}_v \leq 1$.

To assess monotonicity, compute the partial derivatives:

$$\frac{\partial\, \text{SNR}_v}{\partial \delta_v} = \frac{(\delta_v + \eta_v) - \delta_v}{(\delta_v + \eta_v)^2} = \frac{\eta_v}{(\delta_v + \eta_v)^2} > 0,$$

so SNR_v is strictly increasing in δ_v. Moreover,

$$\frac{\partial\, \text{SNR}_v}{\partial \eta_v} = \frac{-\delta_v}{(\delta_v + \eta_v)^2} < 0 \qquad (\text{for } \delta_v > 0),$$

hence SNR_v is strictly decreasing in η_v.

2. Bounds of the composite score. From part (1), we have $0 \leq \text{SNR}_v \leq 1$. By definition, the normalized entropy also satisfies $0 \leq \mathcal{E}_v \leq 1$. The composite score s_v is a convex combination of

these two quantities with weights $\alpha \in [0,1]$ and $(1-\alpha) \in [0,1]$. A convex combination of values in $[0,1]$ also lies in $[0,1]$. Therefore,

$$0 \leq s_v \leq 1.$$

3. Monotonicity of the composite score. The score s_v depends on δ_v through the SNR_v term, and η_v also depends on δ_v via $\lambda_{k+1} = \lambda_k + \delta_v$. We analyze the derivative of s_v with respect to δ_v:

$$\frac{ds_v}{d\delta_v} = \alpha \, \frac{d\,\mathrm{SNR}_v}{d\delta_v}.$$

The derivative of SNR_v with respect to δ_v is

$$\frac{d\,\mathrm{SNR}_v}{d\delta_v} = \frac{\partial\,\mathrm{SNR}_v}{\partial\delta_v} + \frac{\partial\,\mathrm{SNR}_v}{\partial\eta_v}\,\frac{d\eta_v}{d\delta_v}.$$

Since $\eta_v = e^{-t(\lambda_k+\delta_v)}$, we have

$$\frac{d\eta_v}{d\delta_v} = -t\,e^{-t(\lambda_k+\delta_v)} = -t\,\eta_v.$$

Substituting this and the partial derivatives from part (1) yields

$$\frac{d\,\mathrm{SNR}_v}{d\delta_v} = \frac{\eta_v}{(\delta_v+\eta_v)^2} + \frac{-\delta_v}{(\delta_v+\eta_v)^2}\,(-t\eta_v) = \frac{\eta_v\,(1+t\delta_v)}{(\delta_v+\eta_v)^2} \geq 0.$$

Since $\eta_v > 0$, $t > 0$, and $\delta_v \geq 0$, the derivative is nonnegative. As $\alpha \geq 0$, it follows that $\dfrac{ds_v}{d\delta_v} \geq 0$, confirming that s_v is monotonically non-decreasing with respect to the spectral gap δ_v.

E.3 Proof of Lemma 3

Proof. Let S_i denote the event that the sample drawn in the i-th iteration, for $i \in \{1,\ldots,M\}$, consists entirely of inliers. Assuming correspondences are sampled independently and uniformly from a sufficiently large set, the probability that a single randomly chosen correspondence is an inlier is p_{vw}. Hence the probability that all m correspondences in one sample are inliers is

$$\mathbb{P}(S_i) = p_{vw}^m.$$

The complementary event, S_i^c, that the i-th sample contains at least one outlier, has probability

$$\mathbb{P}(S_i^c) = 1 - p_{vw}^m.$$

Since the M iterations are independent, the probability that *all* M samples contain at least one outlier (total failure) is the product of individual failure probabilities:

$$\mathbb{P}\left(S_1^c \cap \cdots \cap S_M^c\right) = \prod_{i=1}^{M} \mathbb{P}(S_i^c) = \left(1 - p_{vw}^m\right)^M.$$

Therefore, the probability that at least one sample consists entirely of inliers (the complement of total failure) is

$$1 - \left(1 - p_{vw}^m\right)^M.$$

To guarantee a success probability of at least $1-\beta$, set

$$1 - \left(1 - p_{vw}^m\right)^M \geq 1 - \beta.$$

Taking natural logarithms yields

$$M\,\log\!\left(1 - p_{vw}^m\right) \leq \log(\beta).$$

Since $p_{vw} \in (0,1]$ and $m \geq 1$, we have $1 - p_{vw}^m \in [0,1)$, so $\log(1-p_{vw}^m) < 0$. Dividing by this negative quantity reverses the inequality, giving the lower bound

$$M \geq \frac{\log(\beta)}{\log\!\left(1 - p_{vw}^m\right)}.$$

This establishes the minimum number of iterations needed to ensure success with the desired probability.

E.4 Proof of Theorem 4

Proof. The proof proceeds in two main steps. First, we invoke the probabilistic guarantee of RANSAC. Second, we apply standard results from matrix perturbation theory to the Orthogonal Procrustes problem solved on the consensus set identified by RANSAC.

By Lemma 3, executing RANSAC for

$$M \geq \frac{\log(\beta)}{\log\left(1 - p_{vw}^{m}\right)}$$

iterations guarantees that, with probability at least $1 - \beta$, at least one of the minimal samples will consist entirely of inliers. The Procrustes solution derived from this all-inlier sample will be a high-quality initial estimate of the true rotation $\mathcal{Q}_{vw}^{*}$. RANSAC then expands this initial estimate to form the largest possible consensus set, C_{vw}, containing all correspondences consistent with this model.

The algorithm's acceptance criterion requires

$$|C_{vw}| \geq d_{\text{adapt}}(s_v, s_w).$$

Let the submatrices of the embeddings corresponding to the points in the consensus set be $A, B \in \mathbb{R}^{|C_{vw}| \times k}$. By assumption, these points are noisy observations of a ground-truth configuration A_0 such that $B_0 = A_0(\mathcal{Q}_{vw}^{*})^{\top}$. The observed matrices can be modeled as

$$A = A_0 + E_A, \qquad B = B_0 + E_B,$$

where E_A and E_B are noise matrices whose entries are i.i.d. sub-Gaussian with parameter σ.

The Procrustes solution $\widehat{\mathcal{Q}}_{vw}$ is found by computing the SVD of the cross-covariance matrix $A^{\top}B = U\Sigma V^{\top}$, yielding $\widehat{\mathcal{Q}}_{vw} = VU^{\top}$. The quality of this estimate depends on how the noise matrices E_A and E_B perturb the singular vectors U and V. Standard matrix–perturbation results (e.g., Davis-Kahan (Davis & Kahan, 1970)) bound the perturbation of singular subspaces; specialized to the rotation error, one obtains that the metric $\|\sin\Theta(\widehat{\mathcal{Q}}_{vw},\, \mathcal{Q}_{vw}^{*})\|_F$ is controlled by the size of the perturbation relative to the singular values of the true cross-covariance $A_0^{\top}B_0$. In particular, a bound of the form

$$\|\sin\Theta(\widehat{\mathcal{Q}}_{vw},\, \mathcal{Q}_{vw}^{*})\|_F \leq O\left(\frac{\|A^{\top}E_B + E_A^{\top}B\|_F}{\sigma_{\min}(A_0^{\top}B_0)}\right)$$

holds. Under the i.i.d. noise model, the expected norm of the perturbation term scales as $\sigma\sqrt{|C_{vw}|}$, while the singular values of $A_0^{\top}B_0$ scale with $|C_{vw}|$. Combining these factors gives the scaling

$$O\left(\frac{\sigma\sqrt{|C_{vw}|}}{|C_{vw}|}\right) = O\left(\frac{\sigma}{\sqrt{|C_{vw}|}}\right).$$

Since a stitch is accepted only if $|C_{vw}| \geq d_{\text{adapt}}(s_v, s_w)$, we substitute this minimum size to obtain the final result: conditioned on RANSAC success,

$$\|\sin\Theta(\widehat{\mathcal{Q}}_{vw},\, \mathcal{Q}_{vw}^{*})\|_F \leq \frac{C\,\sigma}{\sqrt{d_{\text{adapt}}(s_v, s_w)}},$$

for a constant C depending on geometric factors of the underlying configuration. This shows that the adaptive, fidelity-driven evidence requirement enforces a stricter geometric accuracy guarantee for accepted stitches, particularly for those originating from lower-quality data.

While Theorem 4 provides an upper bound on the alignment error for accepted stitches, the following corollary provides a complementary guarantee. It establishes a deterministic condition under which a potential stitch is *guaranteed to be rejected*. This demonstrates that the adaptive threshold actively prunes geometrically unreliable alignments, preventing them from corrupting the reconstruction.

Corollary 10 (Rejection of spurious stitches)**.** *Let an eligible pair of patches (v, w) have an overlap I_{vw} with a true inlier fraction of p_{vw}. The total number of true inliers available in the overlap is*

therefore $p_{vw} \cdot |I_{vw}|$. If this number is less than the minimum required consensus size dictated by the adaptive threshold, i.e.,

$$p_{vw} \cdot |I_{vw}| < d_{adapt}(s_v, s_w),$$

then the RANSAC–Procrustes procedure is guaranteed to reject the stitch between v and w, regardless of the number of iterations performed.

Proof. The RANSAC–Procrustes algorithm returns a consensus set C_{vw}, which by definition contains only correspondences identified as inliers. Let I^*_{vw} denote the (unknown) set of all true inliers in the overlap. This is the largest pool from which any valid consensus can be formed, so

$$|C_{vw}| \leq |I^*_{vw}| = p_{vw}\,|I_{vw}|.$$

The AFR acceptance rule requires the consensus size to meet the adaptive threshold:

$$|C_{vw}| \geq d_{\text{adapt}}(s_v, s_w).$$

By the premise of the corollary,

$$p_{vw}\,|I_{vw}| < d_{\text{adapt}}(s_v, s_w).$$

Combining the inequalities yields

$$|C_{vw}| \leq p_{vw}\,|I_{vw}| < d_{\text{adapt}}(s_v, s_w).$$

Therefore $|C_{vw}| < d_{\text{adapt}}(s_v, s_w)$, so the acceptance criterion can never be satisfied; the stitch is guaranteed to be rejected. This conclusion holds regardless of the number of RANSAC iterations, since no amount of sampling can produce a consensus larger than the total number of true inliers available.

E.5 Intra-island refinement and cross-island voting

This section provides a more detailed mathematical justification for the two post-processing procedures introduced in Stage 3 of the AFR methodology.

Bundle adjustment. For a given reconstructed island G_i, let ε_i be the set of all successfully stitched overlaps. The initial set of relative rotations $\{\hat{\mathcal{Q}}_{vw} \mid (v, w) \in \varepsilon_i\}$ is obtained from Stage 2. Our goal is to find an improved set of rotations $\{\mathcal{Q}^*_{vw}\}$ that minimizes the sum of squared alignment residuals:

$$\{\mathcal{Q}^*_{vw}\} = \underset{\{\mathcal{Q}_{vw} \in SO(k)\}}{\arg\min}\ \Phi_i(\{\mathcal{Q}_{vw}\}) := \sum_{(v,w)\in\varepsilon_i} \|\mathcal{P}'_v - \mathcal{P}'_w \mathcal{Q}_{vw}\|_F^2,$$

where $\mathcal{P}'_v$ and $\mathcal{P}'_w$ are the sub-matrices of the local embeddings corresponding to the nodes in the overlap between patches v and w.

This is a non-convex optimization problem over the product manifold $\mathcal{M} = \times_{j=1}^{|\varepsilon_i|} SO(k)_j$, where $SO(k)$ is the special orthogonal group. Since the objective function Φ_i is smooth on this compact manifold, we can apply standard Riemannian optimization methods (Absil et al., 2007). An iterative method like Riemannian gradient descent generates a sequence of estimates $\{\mathcal{Q}^{(t)}\}$. At each iteration t, the step size η_t is determined via a line search procedure designed to satisfy the *Armijo condition*. This ensures a sufficient decrease in the objective function value:

$$\Phi_i(\{\mathcal{Q}^{(t+1)}\}) \leq \Phi_i(\{\mathcal{Q}^{(t)}\}) - c\eta_t \left\|\operatorname{grad} \Phi_i(\{\mathcal{Q}^{(t)}\})\right\|_F^2,$$

for some $c > 0$. This ensures the sequence of objective values decreases monotonically and is thus guaranteed to converge to a first-order stationary point where the norm of the Riemannian gradient is zero.

Cross-voting for inter-island edges. For any pair of nodes (u, w) belonging to different islands, we formulate a hypothesis test to decide if an edge exists between them ($H_1 : (u, w) \in E$) or not ($H_0 : (u, w) \notin E$). The test statistic is the vote count $C(u, w)$, defined as the number of initial patch overlaps that contained both u and w.

We model the co-occurrence of u and w in any single overlap as a Bernoulli trial. Let the probability of co-occurrence be π_1 if an edge exists (H_1) and π_0 if not (H_0), with the reasonable assumption that $\pi_1 > \pi_0$. If there are m such overlaps in total, the vote count $C(u, w)$ follows a Binomial distribution:

$$C(u, w) \mid H_a \sim \text{Binomial}(m, \pi_a) \quad \text{for } a \in \{0, 1\}.$$

Our decision rule is to infer an edge if $C(u, w)$ exceeds a certain threshold C_0. The reliability of this rule can be analyzed using Chernoff bounds on the tails of the Binomial distribution. The probabilities of Type I and Type II errors are bounded exponentially:

$$\mathbb{P}(\text{False Positive}) = \mathbb{P}(C(u, w) \geq C_0 \mid H_0) \leq \exp\left(-mD_{\text{KL}}(C_0/m \| \pi_0)\right)$$
$$\mathbb{P}(\text{False Negative}) = \mathbb{P}(C(u, w) < C_0 \mid H_1) \leq \exp\left(-mD_{\text{KL}}(C_0/m \| \pi_1)\right)$$

where $D_{\text{KL}}(\cdot \| \cdot)$ is the Kullback-Leibler divergence for Bernoulli variables.

By selecting a threshold C_0 between the expected values $m\pi_0$ and $m\pi_1$, both error probabilities decrease exponentially as the amount of evidence m grows. This strong statistical separation justifies the use of the vote count as a reliable indicator of a true edge.

Instead of a hard threshold, we use a sigmoid function to map the vote count to a probability, which naturally translates stronger evidence (higher counts) into higher confidence:

$$\mathsf{P}_{\text{inter}}(u, w) = \left(1 + e^{-\beta(C(u,w)-C_0)}\right)^{-1}.$$