GAUSSIAN SYNERGY-PRESERVING NODE AGGREGATION FOR RESTRICTED GRAPH

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph summaries are necessarily property-specific: preserving one statistic does not preserve every edge-level or task-dependent quantity. We study a Laplacian-induced Gaussian mutual-information statistic when a set of pairwise nonadjacent nodes is merged. The graph is associated with a possibly degenerate Gaussian representation whose covariance is the weighted graph Laplacian, and all mutual-information calculations are performed on nonsingular proper principal marginals. We characterize a positive model-specific Gaussian synergy gap and derive a closed-form superedge weight that preserves the group-to-target mutual information for a single-external-neighbor motif. For a double-external-neighbor motif, the construction preserves the two marginal mutual-information quantities separately under an explicit solvability condition. Controlled and real-weight experiments verify numerical preservation within the structurally valid regime, while PageRank and spectral results are reported as non-target diagnostics. To sum up, this work provides a new paradigm for edge-weight aggregation in graph summarization with explicit query semantics and a rigorous mathematical interpretation.

1 INTRODUCTION

Undirected graphs are widely used to model relational data in complex systems, social networks, and information diffusion. Nodes represent entities, while weighted edges describe their associations. As networks grow, however, storing, analyzing, and querying the original graph becomes increasingly costly. Graph summarization addresses this problem by merging nodes and edges into a smaller representation while retaining properties relevant to downstream tasks.

Node aggregation is a common approach to graph summarization. When multiple nodes are merged into a summary node, their edges to external nodes must also be combined into summary edges. Determining the aggregated edge weights is therefore a central problem. Existing methods serve different preservation objectives: sum and mean aggregation retain basic statistics, reconstruction-based methods approximate the original adjacency structure, spectral methods preserve global spectral properties, and task-driven methods optimize specific analytical objectives.

Some network-analysis tasks concern the collective association between a node set and a specified target rather than the individual associations considered separately. One motivating setting is group monitoring in a weighted flow network. Edge variables may represent independent perturbations on transport, communication, or exchange channels, while each node records the signed net perturbation incident to it. An analyst may then ask how much observing a candidate node group reduces uncertainty about the net perturbation at a target hub, or rank several candidate groups by this dependence. Such a task is meaningful only when the probabilistic model and the group-to-target query are specified before aggregation.

When a compressed graph is intended to support such a predeclared query, the same dependence score should be recoverable through the summary node and its aggregated edge. Otherwise, aggregation may change both the query value and the ranking of candidate groups. This objective is conditional on the intended use of the summary. If only a fixed list of scalar queries is required, directly caching their values may be simpler; graph aggregation is relevant when a smaller graph representation is also needed for subsequent graph-based processing.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We therefore introduce a property-specific node-aggregation criterion for Gaussian group-to-target dependence queries. Rather than selecting a summary-edge weight by sum or mean aggregation, we derive it by requiring a specified mutual-information quantity to remain unchanged. For a single external neighbor, this quantity is the mutual information between the original node group and the target. For two external neighbors, the construction preserves the two group-to-target marginal mutual-information quantities separately.

Two technical choices are required. First, graph structure and edge weights must be associated with a declared probabilistic representation. We use a zero-mean Gaussian whose covariance is the weighted graph Laplacian. This is a modeling choice with a concrete edge-flow interpretation developed below, rather than a distributional claim about arbitrary observed networks. Second, the summary-edge weights must satisfy the corresponding mutual-information preservation equations under explicitly stated structural, positivity, and nonsingularity assumptions. We study two local settings according to the number of external neighbors: the single-external-neighbor case and the double-external-neighbor case. In the latter case, the preservation targets are the two marginal dependences between the merged node set and the two external neighbors.

In summary, this work makes three contributions: (1) We define a Laplacian-induced Gaussian representation and clarify its probabilistic scope. The full covariance may be singular, while the mutual-information calculations are restricted to nonsingular proper principal marginals. (2) We formulate group-to-target dependence preservation as a node-aggregation criterion. We derive a closed-form rule for a single external neighbor and a conditionally feasible rule that preserves two marginal mutual-information quantities for two external neighbors. (3) We provide numerical implementation checks and model-internal query evaluations within the structurally valid regime. PageRank and spectral measurements are reported as non-target diagnostics and vary across graph families and aggregation orders.

2 Related Work

Graph coarsening is a graph summarization method that reduces graph size by grouping nodes into supernodes and replacing original edges with aggregated superedges Loukas & Vandergheynst (2018). Existing graph coarsening methods can be divided into reconstruction-based and reconstruction-free approaches.

Reconstruction-based methods optimize the coarsening mapping by minimizing discrepancies between the original graph and a reconstructed or lifted graph. Representative objectives include structural reconstruction error LeFevre & Terzi (2010); Riondato et al. (2014), preservation of node features and adjacency properties Kumar et al. (2023), spectral approximation Loukas (2018); Loukas & Vandergheynst (2018); Zhao et al. (2018), graph-neural-network learning of coarse-graph edge weights Cai et al. (2021), and spectrum-preserving coarsening based on Gromov–Wasserstein geometry Chen et al. (2023). These methods provide strong guarantees or empirical fidelity for their respective targets, but their aggregation rules are not designed around collective-dependence queries. Reconstruction-free methods instead construct summaries directly from structural, coding, influence, or task-driven criteria. Examples include attribute- and relation-based interactive aggregation Tian et al. (2008), minimum-description-length summarization Wu et al. (2014), influence-preserving coarsening Purohit et al. (2014), and reinforcement-learning-based task-specific summarization Amiri et al. (2018). These approaches avoid explicit reconstruction objectives and preserve quantities determined by their respective structural or task-specific criteria. Information-oriented graph pooling is also directly relevant. CGIPool maximizes mutual information between input and coarsened graph representations Pang et al. (2021). Its learned representation-level objective differs from the analytic local Gaussian query considered here. Accordingly, our contribution is not the first use of mutual information in graph pooling or coarsening; it is an exact aggregation rule for a specified group-to-target dependence quantity under restricted local motifs.

Existing methods optimize different preservation targets, including reconstruction, spectral similarity, influence, representation-level mutual information, and task-specific performance. The present work adds a narrower target: exact preservation of a Laplacian-induced Gaussian group-to-target dependence query under restricted independent-neighbor motifs. Its aggregation rule is therefore complementary to existing coarsening methods, and its usefulness depends on whether the declared probabilistic model and query class match the intended analysis.

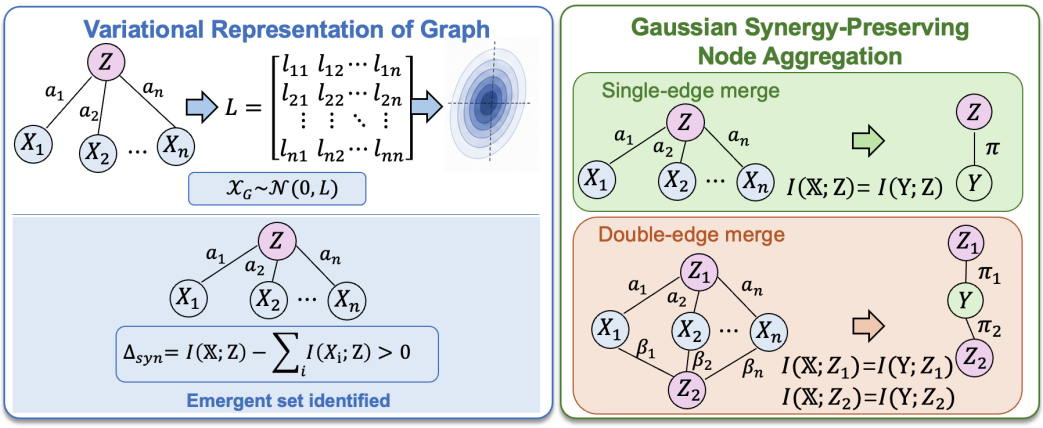

Figure 1: Overview of the proposed Gaussian synergy-preserving graph summarization pipeline. The framework first maps a weighted graph into a Gaussian variational representation through the graph Laplacian, identifies node sets with emergent collective dependence, and then performs single-edge or double-edge aggregation while preserving the target Gaussian mutual-information statistic.

3 Laplacian-Induced Gaussian Representation

As shown in Figure 1, the first step of our framework is to convert the graph topology into a probabilistic representation so that emergence can be quantified by mutual information.

A graph $G = (V, E)$ consists of a node set V and an edge set E. Most conventional aggregation methods mainly operate on direct pairwise edge associations, which makes it difficult for them to describe emergence effects induced by collective dependence. Given that current research on emergence predominantly adopts information-theoretic perspectives for theoretical discussions Hoel (2017); Rosas et al. (2024); Hoel et al. (2013); Williams & Beer (2010), we reconsider graph summarization through mutual information. However, mutual information is defined for random variables, which requires us to map graph topology to a probabilistic representation. Inspired by the mapping between graphs and random variables Tian et al. (2023), we map the topological structure of a graph $G = (V, E)$ to a random vector $\mathcal{X}_G$.

The representation has mean $\mathbb{E}[\mathcal{X}_G] = 0$ and covariance $\mathrm{Cov}(\mathcal{X}_G) = \Sigma$. We define the process of mapping the graph G to $\mathcal{X}_G$ as $\Psi : G \to \mathcal{X}_G$. Then, for each $v_i \in V$, $\mathcal{X}_G(i) = \Psi(v_i)$. According to Shuman et al. (2013), its smoothness is $\mathcal{S}(\Psi) = \mathcal{X}_G^{\mathsf{T}} L \mathcal{X}_G$. Thus, smoothness is defined by the graph Laplacian $L = \mathrm{diag}(\deg(v_1), \ldots, \deg(v_n)) - W$, where $\deg(v_i) = \sum_{j=1}^{n} W_{ij}$. Here, W is the weight matrix of G. We take the weighted graph Laplacian as the covariance by the modeling definition $\Sigma = L$. This choice has a concrete interpretation for edge-flow perturbations. Give each edge an arbitrary orientation, let B be the signed incidence matrix, and let $\Omega = \mathrm{diag}(w_e)$. If independent centered edge perturbations satisfy $f \sim \mathcal{N}(0, \Omega)$, then the corresponding net node perturbations $x = Bf$ satisfy $\mathrm{Cov}(x) = B\Omega B^{\mathsf{T}} = L$. This construction motivates the covariance choice for edge-flow data, but its Gaussian and independence assumptions must be examined before it is applied to a particular domain. It does not imply that arbitrary social, interaction, or co-occurrence networks follow this generative process. For a zero-mean random graph signal with an arbitrary covariance Σ, $\mathbb{E}[\mathcal{X}_G^{\mathsf{T}} L \mathcal{X}_G] = \mathrm{tr}(L\Sigma)$. Under the declared choice $\Sigma = L$, this becomes $\mathrm{tr}(L^2)$. This identity is a consequence of the model rather than a derivation that uniquely determines the covariance.

Once a covariance has been specified, the Gaussian distribution is the maximum-entropy distribution under that covariance constraint. The maximum-entropy principle therefore motivates the Gaussian family after the covariance has been fixed; it does not determine that the covariance must equal the graph Laplacian. Under our modeling choice, $\mathcal{X}_G \sim \mathcal{N}(0, L)$. Because L is positive semidefinite and may be singular, this notation denotes a possibly degenerate Gaussian supported on the range of L. All mutual-information calculations below use finite, nonsingular proper principal marginals.

For a proper jointly Gaussian marginal $(\mathbb{X}, Z)$, let $\mathrm{Cov}(\mathbb{X}) = \Sigma_{\mathbb{X}}$, $\mathrm{Cov}(\mathbb{X}, Z) = p$, and $\mathrm{Var}(Z) = \sigma_Z^2$. Its squared multiple correlation and Gaussian mutual information are

$$R^2_{Z|\mathbb{X}} = \frac{p^\top \Sigma_{\mathbb{X}}^{-1} p}{\sigma_Z^2}, \qquad I(\mathbb{X}; Z) = -\frac{1}{2}\log\Big(1 - R^2_{Z|\mathbb{X}}\Big) = \frac{1}{2}\log\frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z \mid \mathbb{X})}.$$

Thus, the query measures the relative reduction in uncertainty about Z after observing $\mathbb{X}$ under the declared Gaussian model. It does not by itself guarantee preservation of absolute prediction error or arbitrary downstream-task performance.

4 Emergence Node Identification Based on Information Theory

Under the Gaussian graph representation, each node corresponds to one component of the random vector $\mathcal{X}_G$. As shown in Figure 1, we therefore study collective dependence through the mutual information between node variables. Since every subset of a multivariate Gaussian random vector is also Gaussian, the local node variables considered below follow the corresponding marginal Gaussian distributions.

For the declared Gaussian representation, define the Gaussian synergy gap of a node set $\mathbb{X} = \{X_1, X_2, \ldots, X_n\}$ with respect to a target Z as $\Delta_{\mathrm{syn}}(\mathbb{X}; Z) = I(\mathbb{X}; Z) - \sum_{i=1}^n I(X_i; Z)$. A positive value identifies a superadditive dependence pattern under this particular representation. This operational quantity is not a partial-information-decomposition synergy atom, because PID additionally requires a redundancy functional and separates redundant, unique, and synergistic components. We therefore use the term synergy only for this model-specific gap. The following theorem gives a sufficient structural condition for this emergence effect.

Theorem 1 *Let $n \geq 2$, and suppose the queried Gaussian marginal is finite and nonsingular. If the nodes in $\mathbb{X} = \{X_1, X_2, \ldots, X_n\}$ are pairwise nonadjacent and each is connected to Z by a positive-weight edge, then the Gaussian synergy gap is positive:*

$$\Delta_{\mathrm{syn}}(\mathbb{X}; Z) = I(\mathbb{X}; Z) - \sum_{i=1}^n I(X_i; Z) > 0.$$

The proof is provided in Appendix A.1.

5 Gaussian Synergy-Preserving Node Aggregation

After identifying a positive Gaussian synergy gap, we consider local node combinations that satisfy the structural conditions required for dependence-preserving aggregation. According to the number of external neighbors affected by the merge, we divide the aggregation into two cases: single-edge merging and double-edge merging. Under the assumptions of Theorem 1, the node set $\mathbb{X} = \{X_1, X_2, \ldots, X_n\}$ in the following local structures has a positive model-specific Gaussian synergy gap with respect to each corresponding target.

5.1 Single-edge merging

In this case, each $X_i \in \mathbb{X}$ is connected to node Z with weight α_i, and the sum of the weights of the edges connecting X_i to nodes other than Z is r_i. There is no direct edge between any two elements in $\mathbb{X}$, and the sum of the weights of the edges connecting Z to nodes outside $\mathbb{X}$ is o_1. We also require that, except for the shared neighbor Z, the nodes in $\mathbb{X}$ do not share any other external neighbor.

Theorem 2 (Single-target dependence preservation) *Under the stated structural, positivity, and nonsingularity assumptions, after merging all nodes in $\mathbb{X}$ into a summary node Y, there is a unique positive edge weight π satisfying $I(\mathbb{X}; Z) = I(Y; Z)$. This weight is given by:*

$$\pi = \frac{C(\sum_{j=1}^n r_j + o_1)}{2(1-C)} + \frac{\sqrt{C^2(\sum_{j=1}^n r_j + o_1)^2 + 4C(1-C)o_1 \sum_{j=1}^n r_j}}{2(1-C)}, \tag{1}$$

$$C = \sum_{i=1}^n \frac{\alpha_i^2}{(\alpha_i + r_i)(\sum_{j=1}^n \alpha_j + o_1)}. \tag{2}$$

The proof is provided in Appendix A.2.

5.2 DOUBLE-EDGE MERGING

In this case, each $X_i \in \mathbb{X}$ is connected to node Z_1 with weight α_i and to node Z_2 with weight β_i. The sum of the weights of the edges connecting X_i to nodes other than Z_1 and Z_2 is r_i. There is no direct edge between any two elements in $\mathbb{X}$. The sum of the weights of the edges connecting Z_1 to nodes outside $\mathbb{X}$ is o_1, and the sum of the weights of the edges connecting Z_2 to nodes outside $\mathbb{X}$ is o_2. If Z_1 and Z_2 are adjacent, their mutual edge weight is included in both o_1 and o_2, and that edge remains unchanged during the contraction. We also require that, except for the shared external nodes Z_1 and Z_2, the nodes in $\mathbb{X}$ do not share any other external neighbor.

For double-edge merging, the preservation targets are the two marginal Gaussian dependences between the node set $\mathbb{X}$ and the external neighbors Z_1 and Z_2.

Theorem 3 (Two-marginal preservation) *Under the stated structural, positivity, and nonsingularity assumptions, a positive finite pair (π_1, π_2) satisfying $I(\mathbb{X}; Z_1) = I(Y; Z_1)$ and $I(\mathbb{X}; Z_2) = I(Y; Z_2)$ exists if and only if $C_1 + C_2 < 1$. When it exists, the solution is unique and is determined by the following equations. Let*

$$C_1 = \sum_{i=1}^{n} \frac{\alpha_i^2}{(\alpha_i + \beta_i + r_i)(\sum_{j=1}^{n} \alpha_j + o_1)}, C_2 = \sum_{i=1}^{n} \frac{\beta_i^2}{(\alpha_i + \beta_i + r_i)(\sum_{j=1}^{n} \beta_j + o_2)}, \tag{3}$$

$$D = \sum_{j=1}^{n} r_j + \pi_1 + \pi_2. \tag{4}$$

Then

$$\begin{aligned} \pi_1 &= \frac{C_1 D + \sqrt{C_1^2 D^2 + 4 C_1 D o_1}}{2}, \pi_2 = \frac{C_2 D + \sqrt{C_2^2 D^2 + 4 C_2 D o_2}}{2}, \\ D &= \sum_{j=1}^{n} r_j + \frac{C_1 D + \sqrt{C_1^2 D^2 + 4 C_1 D o_1}}{2} + \frac{C_2 D + \sqrt{C_2^2 D^2 + 4 C_2 D o_2}}{2}. \end{aligned} \tag{5}$$

For $C_1 + C_2 < 1$, the scalar equation has a unique finite solution D, which can be computed by bracket expansion followed by bisection. If the condition is not satisfied, the candidate merge is declared infeasible.

The proof is provided in Appendix A.3. When the feasibility condition is violated, the proposed algorithm declares the candidate merge infeasible and does not assign an invalid positive summary weight.

5.3 GRAPH SUMMARIZATION GENERATION PROCESS

The overall procedure iteratively identifies and simplifies candidate subgraphs under the proposed rules. Two candidate sets are maintained: G_S for single-edge merging and G_D for double-edge merging. After each merge, both sets are updated on the current summary graph G', and the process terminates when both are empty, as summarized in Algorithm 1.

6 EXPERIMENTS

We first use controlled simulations as numerical implementation checks for the derived aggregation equations. We then evaluate numerical preservation on observed weighted motifs and use a hub-centered group query to examine query-value and ranking invariance under query-specific contractions. Weighted PageRank and spectral distortion are reported as collateral, *non-target* diagnostics; they are not quantities guaranteed by the proposed aggregation rules. Finally, we examine the empirical feasibility of double-edge merging, sensitivity to graph family and aggregation order, robustness to edge-weight perturbation, and computational cost.

A local merging structure denotes the subgraph induced by the nodes to be merged together with their relevant external neighbors. A structure is considered *valid* only if it satisfies the structural and

Algorithm 1 Generation process of Gaussian-dependence-preserving graph summarization.

Input: Original graph G.
Output: Summary graph G'.

$G' \leftarrow G$, $G_S \leftarrow$ subgraphs in G that support single-edge merging, $G_D \leftarrow$ subgraphs in G that support double-edge merging;
while $G_S \neq \emptyset$ or $G_D \neq \emptyset$ **do**
  if $G_S \neq \emptyset$ **then**
    Select a subgraph from G_S;
    Simplify the selected subgraph in G' based on Theorem 2 and update G_S, G_D;
  end if
  if $G_D \neq \emptyset$ **then**
    Select a subgraph from G_D;
    Simplify the selected subgraph in G' based on Theorem 3 and update G_S, G_D;
  end if
end while
return G';

positivity conditions of the corresponding theorem. Robustness results are therefore evaluated within this structurally valid regime; structurally invalid candidates are rejected by the validity checks.

6.1 Experimental Setup

The real-network experiments use the Zachary Karate Club network Zachary (1977) and the Les Misérables character co-occurrence network Kunegis (2013). The Karate Club graph contains 34 members; its edge weights record the number of interaction contexts observed between pairs of members. The Les Misérables graph contains 77 characters; its edge weights count the chapters in which two characters co-occur. To assess sensitivity to network formation mechanisms, we also generate connected weighted graphs from the Erdős–Rényi Erdős & Rényi (1960), Barabási–Albert Barabási & Albert (1999), and Watts–Strogatz Watts & Strogatz (1998) models. For each family, we generate 20 graphs with 80 nodes and assign independent log-normal edge weights, giving 60 synthetic graph instances in total.

We compare five methods. Proposed uses the Gaussian-dependence-preserving aggregation weight derived in this paper. Sum and Mean respectively sum and average the original edges incident to the shared external node. The *spectral oracle* is a diagnostic control constructed for this study: for the same node set and merge sequence, it uses bounded scalar optimization to select the new edge weight that minimizes the low-frequency normalized-Laplacian error. Heavy-Edge Matching (HEM) follows the classical multilevel coarsening heuristic, pairing each unmatched node with its heaviest unmatched neighbor Karypis & Kumar (1998).

Proposed, Sum, Mean, and the spectral oracle use identical candidate sets and aggregation schedules; only the new edge weights differ. HEM constructs its own matching and contraction map. We therefore match its final number of nodes to that of Proposed, so that compression is treated as a controlled budget rather than as the objective of our method. For local preservation, we report $\left|\rho^2_{\mathrm{before}} - \rho^2_{\mathrm{after}}\right|$. Because Gaussian mutual information is a strictly increasing function of ρ^2, preserving ρ^2 also preserves the corresponding mutual information. For the hub-centered group query, we report mutual-information mean absolute error (MAE), within-hub Spearman rank correlation, and top-10 set overlap. For the non-target end-to-end diagnostics, we report (i) the Spearman correlation between weighted PageRank scores on the original graph and lifted PageRank scores from the summary graph, and (ii) the root mean squared error (RMSE) between nontrivial normalized Laplacian eigenvalues. PageRank Page et al. (1999) is lifted either uniformly or in proportion to the original weighted degree of each constituent node. Spectral error is evaluated at cutoffs $k \in \{5, 10\}$.

6.2 Numerical Verification of the Aggregation Rules

We first verify the closed-form rule for single-edge merging. We generate 500 valid structures, each with five nodes in the emergence node set. For every structure, α_i and r_i are sampled indepen-

Table 1: Median local preservation error on real weighted networks.

Method	Karate Club	Les Misérables
PROPOSED	6.94×10^{-18}	3.47×10^{-18}
SUM	0.00674	0.00571
MEAN	0.0514	0.0226

dently from $\mathcal{U}(1, 9)$, and o_1 is sampled from $\mathcal{U}(1, 30)$. We then compute the new edge weight using PROPOSED, SUM, and MEAN, and compare the squared correlations before and after aggregation.

PROPOSED achieves a median preservation error of 5.55×10^{-17} and a maximum error of 2.22×10^{-16}. The median errors of SUM and MEAN are 0.0270 and 0.302, respectively.

The residual error of PROPOSED is at floating-point precision, confirming that the numerical implementation reproduces the analytical equality in Theorem 2. This is an implementation check of the derived constraint rather than independent evidence for the external validity of the Gaussian model.

We next examine the empirical applicability of the double-edge feasibility condition $C_1 + C_2 < 1$. This experiment evaluates the conditional existence region of the double-edge rule, rather than claiming unconditional solvability for arbitrary local structures. For each $n \in \{2, 5, 10, 25\}$, we generate 500 structures. Under the uniform setting, α_i, β_i, and r_i follow $\mathcal{U}(1, 9)$, while $o_k \sim \mathcal{U}(1, 30)$. All 2,000 structures satisfy the condition.

We then use a more heterogeneous log-normal setting, with $\alpha_i, \beta_i, r_i \sim \mathrm{LogNormal}(0, 1)$ and $o_k \sim \mathrm{LogNormal}(1, 1)$. For $n = 2, 5, 10$, and 25, the condition holds in 99.4%, 97.2%, 91.6%, and 79.2% of the samples, respectively. Thus, larger candidate sets are more likely to approach or cross the feasibility boundary, although most structures generated under this setting still admit a finite positive solution.

For the real networks, we greedily construct one candidate structure for every eligible pair of external nodes. The candidate nodes must be pairwise nonadjacent, their residual neighborhoods—after removing the two shared external nodes—must be disjoint, and all residual weighted degrees must be positive. This procedure yields 12 candidates in Karate Club and 14 in Les Misérables; all satisfy $C_1 + C_2 < 1$.

We next evaluate the proposed rule on structurally valid local motifs drawn from observed weighted networks. For each center node, a deterministic greedy procedure selects at most eight neighbors and checks every structural condition required by the relevant theorem. This yields 10 valid single-edge merging structures in Karate Club and 11 in Les Misérables.

As shown in Table 1, PROPOSED preserves the target statistic to floating-point precision on valid real-weight structures, whereas SUM and MEAN introduce visible deviations.

These results verify numerical preservation for the observed edge weights. They do not establish that the Laplacian-induced Gaussian score has a sociological or literary interpretation for the two networks.

6.3 HUB-CENTERED EMERGENT-GROUP QUERIES

We construct a model-internal query that ranks groups of two to four pairwise nonadjacent neighbors by their Laplacian-induced Gaussian mutual information with a focal node. Valid groups share the focal neighbor, have disjoint residual neighborhoods and positive residual weighted degrees, and therefore satisfy Theorems 1 and 2. Enumeration yields 277 queries around 11 focal nodes in Karate Club and 2,911 queries around 14 focal nodes in Les Misérables.

For each query, we compute the original score and then contract the candidate group separately with PROPOSED, SUM, MEAN, or the spectral oracle. Because groups may overlap, this tests query-specific contraction rather than one reusable summary for all queries. These methods use the same candidates and differ only in the summary-edge weight. HEM is matched to the same node budget; its result is unanswerable when the group is not represented by one summary node or the focal node is contracted. We report absolute error, within-focal Spearman correlation, and top-10 overlap.

Dataset	Method	Queries	Median error	Maximum error	Rank ρ	Top-10 overlap
Karate Club	PROPOSED	277	3.47×10^{-18}	2.78×10^{-17}	1.000	1.000
	SUM	277	0.00265	0.0582	0.926	0.950
	MEAN	277	0.0173	0.0818	0.768	0.950
	Spectral oracle	277	0.0105	0.0919	0.136	0.850
	HEM	277	–	–	–	–
Les Misérables	PROPOSED	2,911	8.67×10^{-19}	4.16×10^{-17}	1.000	1.000
	SUM	2,911	0.00134	0.0838	0.940	1.000
	MEAN	2,911	0.00521	0.0904	0.850	1.000
	Spectral oracle	2,911	0.00414	0.138	0.500	1.000
	HEM	2,911	–	–	–	–

Table 2: Hub-centered emergent-group query results on real weighted networks. A dash indicates that HEM does not retain a summary representation of the pre-registered query.

Table 2 reports the query results. PROPOSED preserves the model-defined values and the associated rankings and top-10 results for each separate contraction, whereas SUM, MEAN, and the spectral oracle alter at least part of the query output; HEM cannot answer the registered queries at the matched budget. This evaluation is model-internal because the rule and score use the same Gaussian representation.

6.4 END-TO-END GRAPH SUMMARIZATION

The end-to-end experiment activates both single-edge and double-edge merging. After each contraction, the current summary graph is updated and the validity conditions of the two merging cases are checked again. The process terminates when no valid structure remains. On Karate Club, the procedure performs eight single-edge and six double-edge merges, reducing the graph from 34 to 18 nodes, corresponding to a compression ratio of 47.1%. On Les Misérables, it performs 13 single-edge and 12 double-edge merges, reducing the graph from 77 to 47 nodes, corresponding to a compression ratio of 39.0%. HEM is constrained to the same final node count on each dataset.

All double-edge structures selected during the end-to-end process satisfy the feasibility condition $C_1 + C_2 < 1$. The largest observed value of $C_1 + C_2$ is 0.102 on Karate Club and 0.099 on Les Misérables.

Across the complete contraction sequences, the maximum absolute errors in the marginal dependence quantities used by the aggregation rules are 7.14×10^{-14} and 4.14×10^{-14}, respectively, which are consistent with the numerical tolerance of the solver.

In Table 3, PR-U and PR-D denote PageRank rank correlation after uniform and weighted-degree-proportional lifting, respectively, while Spec.-10 is the RMSE over the first 10 nontrivial normalized-Laplacian eigenvalues.

Dataset	Method	PR-U	PR-D	Spec.-10
Karate	PROPOSED	0.750	0.951	0.271
	SUM	0.750	0.951	0.274
	MEAN	0.616	0.911	0.279
	Spectral oracle	0.584	0.847	0.178
	HEM	0.622	0.675	0.327
Les Mis.	PROPOSED	0.806	0.893	0.330
	SUM	0.804	0.893	0.330
	MEAN	0.788	0.869	0.328
	Spectral oracle	0.763	0.854	0.239
	HEM	0.901	0.904	0.301

Table 3: End-to-end results with both single-edge and double-edge merging at matched final node counts. Higher is better for PR-U and PR-D; lower is better for Spec.-10.

PROPOSED is designed to preserve the declared local dependence quantities rather than these global graph properties. The PageRank and spectral results therefore characterize collateral distortion under the matched experimental protocol. On Karate Club, PROPOSED obtains the highest PR-D and the lowest Spec.-10 among the non-oracle methods; its PR-U is also higher than those of MEAN, the spectral oracle, and HEM. On Les Misérables, its PR-D is only 0.011 below HEM at the matched 47-node budget, while remaining higher than those of MEAN and the spectral oracle.

These results describe the compression and non-target distortion obtained when both aggregation rules are used sequentially. Activating double-edge merging increases the compression ratio from 44.1% to 47.1% on Karate Club and from 20.8% to 39.0% on Les Misérables.

At the same time, the single-target dependence in each single-edge merge and the two marginal dependence quantities in each double-edge merge are preserved to numerical precision. Thus, the method preserves its declared local dependence quantities, while the non-target PageRank and spectral diagnostics vary with the graph and aggregation procedure.

We repeat the matched-budget comparison on 60 synthetic weighted graphs. The median compression ratios for Erdős–Rényi, Barabási–Albert, and Watts–Strogatz graphs are 62.5%, 76.3%, and 71.3%, respectively. The non-target diagnostics vary across graph families: PROPOSED performs best in PR-D on most Erdős–Rényi and Watts–Strogatz instances, while SUM is stronger on Barabási–Albert PR-D; for spectral error at $k = 10$, PROPOSED performs best on Barabási–Albert graphs, whereas HEM is stronger on Watts–Strogatz graphs. Both HEM and PROPOSED are sensitive to aggregation order. Across 100 random orders, HEM shows substantial variation in PR-D and spectral error. PROPOSED also changes its final compression ratio and non-target diagnostics under different center and candidate orders. These results indicate that order affects global diagnostics, while the target local dependence preservation remains enforced whenever a valid merge is executed.

To assess sensitivity to measurement error, we multiply edge weights by independent log-normal perturbations. Relative to the unperturbed target, the mean drift of the preserved statistic is 0.0204 at noise level $\sigma = 0.2$ and increases to 0.0575 at $\sigma = 0.5$.

The observed drift increases with the perturbation level, quantifying the sensitivity of the model-defined statistic to inaccurate edge weights.

Regarding computational cost, computing C_1 and C_2 for double-edge merging requires one pass over the candidate node set and therefore takes $O(n)$ time. Once these sums are available, each residual evaluation is $O(1)$. Bracket expansion followed by bisection typically requires 38–39 iterations. As n increases from 10 to 30,000, the median solution time grows from 3.2×10^{-5} to 8.7×10^{-5} seconds. Inputs that violate $C_1 + C_2 < 1$ are explicitly reported as infeasible rather than being assigned an invalid positive weight.

These timings characterize only the local weight solver. End-to-end summarization additionally requires candidate discovery, structural validation, conflict resolution, scheduling, and graph updates. Because the complete pipeline is evaluated only on small graphs, the present experiments do not establish large-graph end-to-end scalability.

7 CONCLUSION

We introduced a property-specific node-aggregation rule under a Laplacian-induced Gaussian representation. The framework maps weighted graphs to Gaussian representations and derives aggregation weights by requiring the target mutual-information statistic to remain unchanged after merging. Under the stated structural and feasibility conditions, the proposed rules preserve the target local emergence dependence. The single-edge case admits a closed-form aggregation rule, while the double-edge case is handled through a conditionally feasible solution that rejects structurally or numerically invalid candidates rather than assigning unsupported positive weights. The experiments verify these preservation equations numerically on structurally valid motifs and evaluate query-specific ranking consistency. PageRank and spectral measurements are reported as non-target diagnostics and vary across graphs and aggregation procedure. Overall, this work provides a novel probabilistic and query-aligned criterion for edge-weight aggregation in graph summarization.

AI USE STATEMENT

Generative AI tools, including OpenAI Codex, were used to polish the manuscript's writing, assist with literature search and review, support code development and consistency checking, and create the paper teaser and illustrative icons in the pipeline diagram. AI-generated visual content is limited to these illustrative elements and does not include experimental data, plots, or dataset-derived examples. All AI-assisted text, code, references, and visual elements were reviewed and verified by the authors, who take full responsibility for the content, reported results, and integrity of the manuscript.

Ethics Statement

This work studies graph summarization using synthetic graphs and two publicly available weighted benchmark networks. It does not involve new human-subject data collection or access to private user data. Nevertheless, graphs can encode sensitive relationships, and aggregation may obscure individual nodes, minority structures, or relationships that matter in a particular application. Moreover, the Laplacian-induced Gaussian mutual-information statistic used here is model-specific and should not be interpreted as establishing causal, sociological, or behavioral conclusions about the entities represented by a network. Applications to person-level or otherwise sensitive relational data should follow applicable data-governance and privacy requirements, assess possible representational harms, and evaluate downstream distortion before deployment.

Reproducibility Statement

The main paper specifies the Laplacian-induced Gaussian representation, structural assumptions, preservation objectives, aggregation weights, feasibility condition, and complete summarization algorithm. Appendix A provides proofs of the three theoretical results, while Section 6 reports the synthetic-graph generation procedure, public benchmark datasets, baselines, evaluation metrics, and experimental protocols. The experiments use no restricted or proprietary data. The implementation and experiment scripts will be released to support reproduction of the reported numerical checks and evaluations.

References

Sorour E. Amiri, Bijaya Adhikari, Aditya Bharadwaj, and B. Aditya Prakash. Netgist: Learning to generate task-based network summaries. In *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 857–862, 2018. doi: 10.1109/ICDM.2018.00101.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999. doi: 10.1126/science.286.5439.509.

Chen Cai, Dingkang Wang, and Yusu Wang. Graph coarsening with neural networks, 2021. URL https://arxiv.org/abs/2102.01350.

Yifan Chen, Rentian Yao, Yun Yang, and Jie Chen. A gromov-Wasserstein geometric view of spectrum-preserving graph coarsening. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 5257–5281. PMLR, 2023. URL https://proceedings.mlr.press/v202/chen23ak.html.

Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

Erik P. Hoel. When the map is better than the territory. *Entropy*, 19(5):188, 2017. doi: 10.3390/e19050188.

Erik P. Hoel, Larissa Albantakis, and Giulio Tononi. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*, 110(49):19790–19795, 2013. doi: 10.1073/pnas.1314922110. URL https://www.pnas.org/doi/abs/10.1073/pnas.1314922110.

George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20(1):359–392, 1998. doi: 10.1137/S1064827595287997.

Manoj Kumar, Anurag Sharma, Shashwat Saxena, and Sandeep Kumar. Featured graph coarsening with similarity guarantees. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Jérôme Kunegis. Konect: The koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web Companion*, pp. 1343–1350, New York, NY, USA, 2013. Association for Computing Machinery. doi: 10.1145/2487788.2488173.

Kristen LeFevre and Evimaria Terzi. Grass: Graph structure summarization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 454–465, 2010. doi: 10.1137/1.9781611972801.40.

Andreas Loukas. Graph reduction with spectral and cut guarantees, 2018. URL https://arxiv.org/abs/1808.10650.

Andreas Loukas and Pierre Vandergheynst. Spectrally approximating large graphs with smaller graphs, 2018. URL https://arxiv.org/abs/1802.07510.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999. URL https://ilpubs.stanford.edu:8090/422/.

Yunsheng Pang, Yunxiang Zhao, and Dongsheng Li. Graph pooling via coarsened graph infomax. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2177–2181. ACM, 2021. doi: 10.1145/3404835.3463074.

Manish Purohit, B. Aditya Prakash, Chanhyun Kang, Yao Zhang, and V.S. Subrahmanian. Fast influence-based coarsening for large networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pp. 1296–1305, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329569. doi: 10.1145/2623330.2623701. URL https://doi.org/10.1145/2623330.2623701.

Matteo Riondato, David García-Soriano, and Francesco Bonchi. Graph summarization with quality guarantees. In *2014 IEEE International Conference on Data Mining*, pp. 947–952, 2014. doi: 10.1109/ICDM.2014.56.

Fernando E. Rosas, Bernhard C. Geiger, Andrea I Luppi, Anil K. Seth, Daniel Polani, Michael Gastpar, and Pedro A. M. Mediano. Software in the natural world: A computational approach to hierarchical emergence, 2024. URL https://arxiv.org/abs/2402.09090.

David I Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013. doi: 10.1109/MSP.2012.2235192.

Yang Tian, Hedong Hou, Guangzheng Xu, Ziyang Zhang, and Pei Sun. Network comparison via encoding, decoding, and causality, 2023. URL https://arxiv.org/abs/2207.06606.

Yuanyuan Tian, Richard A. Hankins, and Jignesh M. Patel. Efficient aggregation for graph summarization. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pp. 567–580, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605581026. doi: 10.1145/1376616.1376675. URL https://doi.org/10.1145/1376616.1376675.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998. doi: 10.1038/30918.

Paul L. Williams and Randall D. Beer. Nonnegative decomposition of multivariate information, 2010. URL https://arxiv.org/abs/1004.2515.

Ye Wu, Zhinong Zhong, Wei Xiong, and Ning Jing. Graph summarization for attributed graphs. In *2014 International Conference on Information Science, Electronics and Electrical Engineering*, volume 1, pp. 503–507, 2014. doi: 10.1109/InfoSEEE.2014.6948163.

Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977. doi: 10.1086/jar.33.4.3629752.

Zhiqiang Zhao, Yongyu Wang, and Zhuo Feng. Nearly-linear time spectral graph reduction for scalable graph partitioning and data visualization, 2018. URL https://arxiv.org/abs/1812.08942.

A PROOFS

A.1 PROOF FOR THEOREM 1

Let $n \geq 2$, and assume that the queried marginal is finite and positive definite. Suppose the nodes $X_1, \ldots, X_n$ are pairwise nonadjacent and are connected to Z by strictly positive edge weights.

Since the original graph can be variableized into a high-dimensional Gaussian distribution, the joint distribution of $\mathbb{X}$ and Z, as the marginal distribution of the original distribution, will also follow a high-dimensional Gaussian distribution. We assume that $\mathbb{X}, Z \sim N(0, \Sigma_{(\mathbb{X},Z)})$, $\Sigma_{\mathbb{X}}$ represents the covariance matrix composed of the nodes contained in $\mathbb{X}$ and $|\mathbb{X}| = n$, and the variance of the component corresponding to the node Z is σ_Z^2.

In this situation, the covariance matrix $\Sigma_{(\mathbb{X},Z)}$ can be partitioned into the following form, that is:

$$\Sigma_{(\mathbb{X},Z)} = \begin{bmatrix} \Sigma_{\mathbb{X}} & P \\ P^T & \sigma_Z^2 \end{bmatrix}. \tag{6}$$

where $P = [\rho_1\sigma_1\sigma_Z, \rho_2\sigma_2\sigma_Z, \cdots, \rho_n\sigma_n\sigma_Z]^T$, ρ_i represents the correlation coefficient between the i-th element in the node set $\mathbb{X}$ and the node Z, and σ_i represents the standard deviation of the i-th element in the node set $\mathbb{X}$.

Since there are no direct edges between any two nodes contained in the node set $\mathbb{X}$, and $\Sigma_{\mathbb{X}}$ is the covariance matrix of the node set $\mathbb{X}$, $\Sigma_{\mathbb{X}}$ is a diagonal matrix and must be invertible. Thus, we perform the following elementary transformation on $\Sigma_{(\mathbb{X},Z)}$, and denote the resulting matrix as $\Sigma'_{(\mathbb{X},Z)}$:

$$\begin{aligned} \Sigma'_{(\mathbb{X},Z)} &= \begin{bmatrix} I_n & 0 \\ -P^T\Sigma_{\mathbb{X}}^{-1} & 1 \end{bmatrix} \times \Sigma_{(\mathbb{X},Z)} \times \begin{bmatrix} I_n & -\Sigma_{\mathbb{X}}^{-1}P \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{\mathbb{X}} & 0 \\ 0 & \sigma_Z^2 - P^T\Sigma_{\mathbb{X}}^{-1}P \end{bmatrix}. \end{aligned} \tag{7}$$

In this way, $\det(\Sigma_{(\mathbb{X},Z)}) = \det(\Sigma'_{(\mathbb{X},Z)}) = \det(\Sigma_{\mathbb{X}}) \times \det(\sigma_Z^2 - P^T\Sigma_{\mathbb{X}}^{-1}P)$. In fact, we have:

$$P^T\Sigma_{\mathbb{X}}^{-1}P = \sigma_Z^2 \times \sum_{i=1}^{n} \rho_i^2. \tag{8}$$

Therefore, for $I(\mathbb{X}; Z)$, according to the relationship between mutual information and entropy, we have:

$$\begin{aligned} I(\mathbb{X}; Z) &= H(\mathbb{X}) + H(Z) - H(\mathbb{X}, Z) \\ &= \frac{1}{2}\log\frac{\det(\Sigma_{\mathbb{X}}) \times \sigma_Z^2}{\det(\Sigma_{(\mathbb{X},Z)})} \\ &= \frac{1}{2}\log\frac{\det(\Sigma_{\mathbb{X}}) \times \sigma_Z^2}{\det(\Sigma_{\mathbb{X}}) \times \det(\sigma_Z^2 - \sigma_Z^2 \times \sum_{i=1}^{n} \rho_i^2)} \\ &= \frac{1}{2}\log\frac{1}{1 - \sum_{i=1}^{n} \rho_i^2}. \end{aligned} \tag{9}$$

Since $\Sigma_{(\mathbb{X},Z)}$ is related to the Laplacian matrix of the original graph, according to the properties of the Laplacian matrix, we have

$$\rho_i^2 = \frac{\alpha_i^2}{(\alpha_i + r_i) \times (\sum_{j=1}^{n} \alpha_j + o)}. \tag{10}$$

where α_i is the weight of the edge connecting the node X_i and the node Z, r_i is the sum of the weights of the edges connecting the node X_i and the nodes other than the node Z, and o is the sum

of the weights of the edges connecting the node Z and the nodes other than those contained in $\mathbb{X}$. Thus, we have

$$1-\sum_{i=1}^{n}\rho_i^2 = 1-\sum_{i=1}^{n}\frac{\alpha_i^2}{(\alpha_i+r_i)\times(\sum_{j=1}^{n}\alpha_j+o)} \\ > 1-\sum_{i=1}^{n}\frac{\alpha_i^2}{\alpha_i\times\sum_{j=1}^{n}\alpha_j} = 1-\frac{\sum_{i=1}^{n}\alpha_i}{\sum_{j=1}^{n}\alpha_j} = 0. \tag{11}$$

For $I(X_i;Z)$, according to the relationship between the mutual information of Gaussian data and the correlation coefficient, we have:

$$I(X_i;Z) = -\frac{1}{2}\log(1-\rho_i^2). \tag{12}$$

In this way, for $\sum_{i=1}^{n} I(X_i;Z)$, we have

$$\sum_{i=1}^{n} I(X_i;Z) = -\frac{1}{2}\sum_{i=1}^{n}\log(1-\rho_i^2) \\ = \frac{1}{2}\log\frac{1}{\prod_{i=1}^{n}(1-\rho_i^2)}. \tag{13}$$

Therefore, to compare the magnitudes of $I(\mathbb{X};Z)$ and $\sum_{i=1}^{n} I(X_i;Z)$, in fact, we need to compare the magnitudes of $1-\sum_{i=1}^{n}\rho_i^2$ and $\prod_{i=1}^{n}(1-\rho_i^2)$. Next, we prove by mathematical induction that when $k\geq 2$, for $0<\rho_i^2<1$ and $1-\sum_{i=1}^{n}\rho_i^2>0$, we always have $\prod_{i=1}^{n}(1-\rho_i^2) > 1-\sum_{i=1}^{n}\rho_i^2$.

When $n=2$, we have

$$\prod_{i=1}^{2}(1-\rho_i^2) = (1-\rho_1^2)(1-\rho_2^2) = 1-\rho_1^2-\rho_2^2+\rho_1^2\rho_2^2 \\ > 1-\sum_{i=1}^{2}\rho_i^2. \tag{14}$$

$\prod_{i=1}^{n}(1-\rho_i^2) > 1-\sum_{i=1}^{n}\rho_i^2$ holds.

Assume that for any $n=k(k\geq 2)$ and $0<\rho_i^2<1$, $\prod_{i=1}^{n}(1-\rho_i^2) > 1-\sum_{i=1}^{n}\rho_i^2$ always holds. That is, assume

$$\prod_{i=1}^{k}(1-\rho_i^2) > 1-\sum_{i=1}^{k}\rho_i^2. \tag{15}$$

Then for the $(k+1)$-th variable added to the node set $\mathbb{X}$, we have

$$(1-\rho_{k+1}^2)\prod_{i=1}^{k}(1-\rho_i^2) > (1-\rho_{k+1}^2)(1-\sum_{i=1}^{k}\rho_i^2). \tag{16}$$

When $1-\sum_{i=1}^{k}\rho_i^2>0$, at this time $0<\rho_{k+1}^2<1$ and $0<\sum_{i=1}^{k}\rho_i^2<1$. From the case of $k=2$, we have

$$(1-\rho_{k+1}^2)(1-\sum_{i=1}^{k}\rho_i^2) > 1-\sum_{i=1}^{k+1}\rho_i^2. \tag{17}$$

Combined with 16, we have

$$(1-\rho_{k+1}^2)\prod_{i=1}^{k}(1-\rho_i^2) > (1-\rho_{k+1}^2)(1-\sum_{i=1}^{k}\rho_i^2) \\ > 1-\sum_{i=1}^{k+1}\rho_i^2. \tag{18}$$

That is

$$\prod_{i=1}^{k+1}(1-\rho_i^2) > 1 - \sum_{i=1}^{k+1}\rho_i^2. \tag{19}$$

Therefore, when $n = k+1$ with $k \geq 2$, $\prod_{i=1}^{k+1}(1-\rho_i^2) > 1 - \sum_{i=1}^{k+1}\rho_i^2$ also holds.

In conclusion, when $n \geq 2$, for $0 < \rho_i^2 < 1$, we always have $\prod_{i=1}^{n}(1-\rho_i^2) > 1 - \sum_{i=1}^{n}\rho_i^2$. Therefore, we always have $\frac{1}{\prod_{i=1}^{n}(1-\rho_i^2)} < \frac{1}{1-\sum_{i=1}^{n}\rho_i^2}$, that is $\sum_{i=1}^{n} I(X_i;Z) < I(\mathbb{X};Z)$. That is, when there are no direct edges between any two nodes contained in the node set $\mathbb{X} = \{X_1, X_2, \cdots, X_n\}$, but all are directly connected to the node Z, there must be an emergence-effect of the node set $\mathbb{X}$ with respect to the node Z.

A.2 Proof for Theorem 2

In the case of single-edge merging, $\mathcal{X}_G$ has a covariance matrix as follows:

$$\Sigma_{\mathcal{X}_G} = \begin{bmatrix} \mathrm{diag}(\alpha_1 + r_1, \ldots, \alpha_n + r_n) & -\boldsymbol{\alpha} \\ -\boldsymbol{\alpha}^\top & \sum_{i=1}^{n}\alpha_i + o_1 \end{bmatrix},$$

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top.$$

In this way, we have the correlation coefficient between X_i and Z as $\rho_i = -\frac{\alpha_i}{\sqrt{\alpha_i + r_i}\sqrt{\sum_{j=1}^{n}\alpha_j + o_1}}$.

All the nodes contained in $\mathbb{X}$ are merged to form the node Y. At this time, $\mathcal{X}_{G'}$ has a covariance matrix as:

$$\Sigma_{\mathcal{X}_{G'}} = \begin{bmatrix} \sum_{i=1}^{n} r_i + \pi & -\pi \\ -\pi & o_1 + \pi \end{bmatrix}.$$

In this way, we have the correlation coefficient between Y and Z as $\rho_y = -\frac{\pi}{\sqrt{\sum_{j=1}^{n} r_j + \pi}\sqrt{o_1 + \pi}}$.

According to 9 and 12, in order to keep the connection between the nodes before and after summarization unchanged, it is necessary to have $\sum_{i=1}^{n}\rho_i^2 = \rho_y^2$, namely:

$$\sum_i \frac{\alpha_i^2}{(\alpha_i + r_i)(\sum_{j=1}^{n}\alpha_j + o_1)} = \frac{\pi^2}{(\sum_{j=1}^{n} r_j + \pi)(o_1 + \pi)}. \tag{20}$$

Let $C = \sum_{i=1}^{n}\frac{\alpha_i^2}{(\alpha_i + r_i)(\sum_{j=1}^{n}\alpha_j + o_1)}$, 20 can be simplified to:

$$\frac{\pi^2}{(\pi + \sum_{j=1}^{n} r_j)(\pi + o_1)} = C, \tag{21}$$

$$(1-C)\pi^2 - C(\sum_{j=1}^{n} r_j + o_1)\pi - C o_1 \sum_{j=1}^{n} r_j = 0. \tag{22}$$

According to the quadratic formula for solving quadratic equations, we have:

$$\pi = \frac{C(\sum_{j=1}^{n} r_j + o_1)}{2(1-C)} \pm \frac{\sqrt{C^2(\sum_{j=1}^{n} r_j + o_1)^2 + 4C(1-C)o_1\sum_{j=1}^{n} r_j}}{2(1-C)}. \tag{23}$$

Since the physical meaning of π is the edge weight in the graph, we require that $\pi > 0$. First, we check the range of values of C.

$$\begin{aligned} C &= \sum_i \frac{\alpha_i^2}{(\alpha_i + r_i)(\sum_j \alpha_j + o_1)} \\ &= \frac{1}{\sum_j \alpha_j + o_1}\sum_i \frac{\alpha_i^2}{\alpha_i + r_i}. \end{aligned} \tag{24}$$

Since $r_i, o_1 > 0$, we have:

$$\sum_i \frac{\alpha_i^2}{\alpha_i + r_i} < \sum_i \frac{\alpha_i^2}{\alpha_i} = \sum_i \alpha_i, \frac{1}{\sum_j \alpha_j + o_1} < \frac{1}{\sum_j \alpha_j}. \tag{25}$$

Therefore, we have:

$$\frac{1}{\sum_j \alpha_j + o_1} \sum_i \frac{\alpha_i^2}{\alpha_i + r_i} < \frac{\sum_i \alpha_i}{\sum_j \alpha_j} = 1. \tag{26}$$

Therefore, $C < 1$ always holds. π can only take the positive root, and the unique solution for π is:

$$\begin{aligned}\pi =& \frac{C(\sum_{j=1}^n r_j + o_1)}{2(1-C)} + \\ & \frac{\sqrt{C^2(\sum_{j=1}^n r_j + o_1)^2 + 4C(1-C)o_1 \sum_{j=1}^n r_j}}{2(1-C)}.\end{aligned} \tag{27}$$

A.3 Proof for Theorem 3

In the case of double-edge merging, $\mathcal{X}_G$ has a covariance matrix as follows. Let $\gamma \geq 0$ denote the weight of the edge between Z_1 and Z_2, with $\gamma = 0$ when they are nonadjacent. This weight is included in both o_1 and o_2 and remains unchanged during contraction:

$$\begin{aligned}&\Sigma_{\mathcal{X}_G} = \\ &\begin{bmatrix} \mathrm{diag}(\alpha_i + \beta_i + r_i)_{i=1}^n & -\boldsymbol{\alpha} & -\boldsymbol{\beta} \\ -\boldsymbol{\alpha}^\top & \sum_{i=1}^n \alpha_i + o_1 & -\gamma \\ -\boldsymbol{\beta}^\top & -\gamma & \sum_{i=1}^n \beta_i + o_2 \end{bmatrix} \\ &\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top,\ \boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\top.\end{aligned}$$

In this way, we have the correlation coefficient between X_i and Z_1 as $\rho_1^i = -\frac{\alpha_i}{\sqrt{\alpha_i+\beta_i+r_i}\sqrt{\sum_{j=1}^n \alpha_j + o_1}}$, the correlation coefficient between X_i and Z_2 as $\rho_2^i = -\frac{\beta_i}{\sqrt{\alpha_i+\beta_i+r_i}\sqrt{\sum_{j=1}^n \beta_j + o_2}}$.

All the nodes contained in $\mathbb{X}$ are merged to form the node Y. At this time, $\mathcal{X}_{G'}$ has a covariance matrix as:

$$\Sigma_{\mathcal{X}_{G'}} = \begin{bmatrix} \sum_{i=1}^n r_i + \pi_1 + \pi_2 & -\pi_1 & -\pi_2 \\ -\pi_1 & o_1 + \pi_1 & -\gamma \\ -\pi_2 & -\gamma & o_2 + \pi_2 \end{bmatrix}.$$

The correlation coefficient between Y and Z_1 is $\rho_1^y = -\frac{\pi_1}{\sqrt{\sum_{j=1}^n r_j + \pi_1 + \pi_2}\sqrt{o_1 + \pi_1}}$, and the correlation coefficient between Y and Z_2 is $\rho_2^y = -\frac{\pi_2}{\sqrt{\sum_{j=1}^n r_j + \pi_1 + \pi_2}\sqrt{o_2 + \pi_2}}$. Because the edge of weight γ is unchanged and the targets are the two marginal mutual informations, preservation requires

$$I(\mathbb{X}; Z_1) = I(Y; Z_1), \qquad I(\mathbb{X}; Z_2) = I(Y; Z_2),$$

or equivalently $\sum_{i=1}^n {\rho_1^i}^2 = {\rho_1^y}^2$ and $\sum_{i=1}^n {\rho_2^i}^2 = {\rho_2^y}^2$, namely:

$$\begin{cases} \sum_{i=1}^n \frac{\alpha_i^2}{(\alpha_i+\beta_i+r_i)(\sum_{j=1}^n \alpha_j + o_1)} = \frac{\pi_1^2}{(\sum_{j=1}^n r_j + \pi_1 + \pi_2)(o_1 + \pi_1)}, \\ \sum_{i=1}^n \frac{\beta_i^2}{(\alpha_i+\beta_i+r_i)(\sum_{j=1}^n \beta_j + o_2)} = \frac{\pi_2^2}{(\sum_{j=1}^n r_j + \pi_1 + \pi_2)(o_2 + \pi_2)}. \end{cases} \tag{28}$$

Let $C_1 = \sum_{i=1}^n \frac{\alpha_i^2}{(\alpha_i+\beta_i+r_i)(\sum_{j=1}^n \alpha_j + o_1)}, C_2 = \sum_{i=1}^n \frac{\beta_i^2}{(\alpha_i+\beta_i+r_i)(\sum_{j=1}^n \beta_j + o_2)}, D = \sum_{j=1}^n r_j + \pi_1 + \pi_2$, 28 can be simplified to:

$$\begin{cases} C_1 = \frac{\pi_1^2}{D(o_1 + \pi_1)}, \\ C_2 = \frac{\pi_2^2}{D(o_2 + \pi_2)}. \end{cases} \tag{29}$$

Therefore, we have:

$$\begin{cases} \pi_1^2 - C_1 D\pi_1 - C_1 Do_1 = 0, \\ \pi_2^2 - C_2 D\pi_2 - C_2 Do_2 = 0. \end{cases} \tag{30}$$

According to the quadratic formula for solving quadratic equations, we can obtain:

$$\begin{cases} \pi_1 = \frac{C_1 D \pm \sqrt{C_1^2 D^2 + 4C_1 Do_1}}{2}, \\ \pi_2 = \frac{C_2 D \pm \sqrt{C_2^2 D^2 + 4C_2 Do_2}}{2}. \end{cases} \tag{31}$$

Since $\pi_1, \pi_2 > 0$, we retain the positive roots here. Substituting π_1 and π_2 into $D = \sum_{j=1}^{n} r_j + \pi_1 + \pi_2$, we get:

$$D = \sum_{j=1}^{n} r_j + \frac{C_1 D + \sqrt{C_1^2 D^2 + 4C_1 Do_1}}{2} + \frac{C_2 D + \sqrt{C_2^2 D^2 + 4C_2 Do_2}}{2}. \tag{32}$$

Let $R = \sum_{j=1}^{n} r_j$ and let $g_k(D)$ denote the positive-root expression for π_k in (31). Dividing (32) by D gives

$$1 = \frac{R}{D} + \frac{g_1(D)}{D} + \frac{g_2(D)}{D}.$$

The right-hand side is continuous and strictly decreasing on $D > 0$, and its limit as $D \to \infty$ is $C_1 + C_2$. Under the stated positivity assumptions, it exceeds one for sufficiently small positive D. Hence a positive finite solution exists if and only if $C_1 + C_2 < 1$, and strict monotonicity makes that solution unique. It can be computed by bracket expansion followed by bisection, after which π_1 and π_2 are obtained from the positive-root expressions. Thus, when the solution exists, it is unique and satisfies

$$I(\mathbb{X}; Z_1) = I(Y; Z_1), \qquad I(\mathbb{X}; Z_2) = I(Y; Z_2).$$