WHEN MISSINGNESS FOLLOWS THE GRAPH: BEYOND FEATURE PROPAGATION FOR MISSING FEATURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph Neural Networks (GNNs) commonly assume that node features are fully observed, despite missing features being pervasive in real-world applications. By leveraging the dependence between node features and graph structure, existing graph-specific methods handle missingness by propagating observed feature values along graph edges. However, these methods have mainly been evaluated in overly optimistic settings, where feature entries are removed independently and uniformly across the graph. This overlooks realistic scenarios in which missingness propagates among connected nodes, as can occur during localized sensing failures in sensor networks. We formalize this complementary notion as *graph-structured mask dependence*, where the joint distribution of missingness indicators is shaped by the graph, and introduce a controlled mechanism that instantiates this setting. We study propagation-based methods theoretically and empirically in this setting, showing that their effectiveness changes substantially compared with the uniformly random regime. Motivated by this limitation, we propose PEMix, designed to remain effective under challenging graph-structured missingness. Rather than propagating observed values, PEMix uses global structural representations to condition a probabilistic model of incomplete node features before message passing. Experiments under controlled graph-structured missingness and naturally occurring missingness show that PEMix remains robust across complex missingness regimes, with particularly strong performance as missingness becomes severe.

1 INTRODUCTION

Graph Neural Networks (GNNs) are commonly designed under the assumption that node features are fully observed (Kipf & Welling, 2017; Hamilton et al., 2017; Velickovic et al., 2018). In practice, this assumption is often violated. Sensors may fail to record part of their measurements (Braem et al., 2024; Faizin et al., 2019), while communication, acquisition, or infrastructure failures can also lead to missing observations.

Naive approaches rely on simple statistics to fill missing data, treating the feature matrix as tabular (Ferrini et al., 2026; Taguchi et al., 2021). Graph-specific methods instead leverage the graph topology to handle missingness more effectively. Most existing methods estimate missing features by propagating the observed values while minimizing a Dirichlet energy (Rossi et al., 2022). Methods such as PCFI and FISF extend this principle with confidence-based and iterative propagation schemes (Um et al., 2023; 2025) and achieve strong empirical performance on downstream node-level tasks.

However, performance is mostly evaluated on an oversimplified family of missingness patterns. The standard protocol removes feature entries independently under uniform Missing Completely At Random (MCAR) (Rubin, 1976). More recent work has considered feature- and label-dependent mechanisms, including Missing Not At Random (MNAR) settings (Ferrini et al., 2026), but these protocols still focus on how missingness depends on feature values or labels. They leave unexplored a different aspect of the problem: how missingness is conditioned by topology. The latter is particularly relevant in applications where failures propagate locally. In sensor networks or smart grids, physical or communication faults often cause entire connected regions of nodes to drop their measured features simultaneously (Marisca et al., 2024). Consequently, missingness indicators violate standard spatial independence assumptions, generating correlated missing patterns across adjacent nodes. Thus, the same global missingness rate can lead to very different levels of difficulty for graph-based

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.

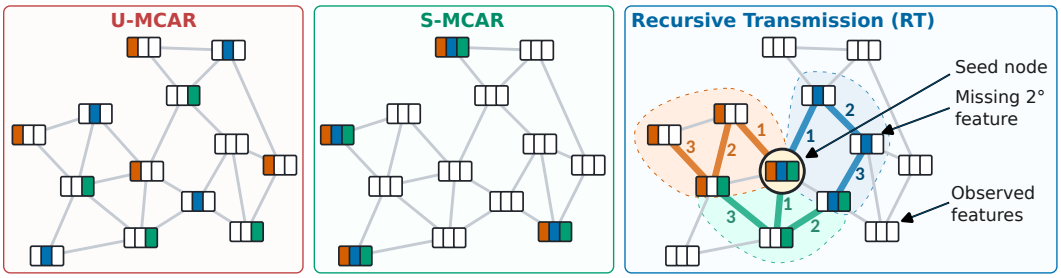


Figure 1: **Conceptual comparison of missingness mechanisms.** *Left (U-MCAR):* Feature entries are dropped independently and uniformly across the graph. *Middle (S-MCAR):* Entire feature vectors are missing for randomly sampled nodes, independent of connectivity. *Right (Recursive Transmission, RT):* Missingness originates at a shared seed node and expands recursively along graph edges, forming spatially correlated missing regions that challenge propagation-based recovery.

methods depending on how missing entries are distributed over the graph. These patterns challenge propagation-based methods: when missing-feature nodes cluster together, there is too little observed information to propagate. We analyze this limitation theoretically and empirically and we propose PEMix, designed to remain effective under challenging graph-structured missingness.

Our contributions are as follows.

1. We distinguish the classical dependence of the missingness mask on the data from the dependence structure within the mask itself, formalize *graph-structured mask dependence*, and introduce Recursive Transmission (RT), a controlled graph-structured mechanism.
2. We establish a boundary bottleneck for standard feature propagation and show empirically that graph-localized missingness challenges more sophisticated propagation methods.
3. We propose PEMix to handle clustered missingness where propagation-based approaches fail. It conditions a probabilistic feature model on global structural representations prior to message passing, incorporating graph topology directly without relying on value propagation.
4. Across controlled graph-structured missingness and real-world datasets, we show that PEMix improves robustness over propagation baselines, achieving particularly strong gains as missingness severity increases.

2 Learning with Missing Node Features

In this section, we follow the notation of Ferrini et al. (2026). We consider a connected attributed graph $G = (V, E, \mathbf{X}, \mathbf{Y})$, where $V = \{1, \dots, n\}$ is the set of nodes, $E \subseteq V \times V$ is the set of edges represented by the adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$, $\mathbf{X} \in \mathbb{R}^{n \times d}$ is the node feature matrix, with X_{ij} denoting feature j of node i, and $\mathbf{Y} \in \{1, \dots, K\}^n$ is the vector of node labels. Throughout the paper, $\mathbf{A}$ is a random variable in the probabilistic formulation and is fully observed. We condition on its realized value throughout learning and inference.

Let $\mathbf{M} \in \{0,1\}^{n \times d}$ denote the missingness mask, where $M_{ij} = 1$ iff X_{ij} is unobserved. We write $\mathbf{X}_{\text{obs}}$ and $\mathbf{X}_{\text{mis}}$ for the observed and missing entries induced by $\mathbf{M}$, respectively. The learner receives $(\mathbf{A}, \mathbf{X}_{\text{obs}}, \mathbf{M})$ together with the labels of the training nodes.

The complete-data distribution can be factorized as

$$p_{\theta,\gamma,\lambda}(\mathbf{X}, \mathbf{Y}, \mathbf{M} \mid \mathbf{A}) = p_\theta(\mathbf{X} \mid \mathbf{A})\, p_\gamma(\mathbf{Y} \mid \mathbf{X}, \mathbf{A})\, p_\lambda(\mathbf{M} \mid \mathbf{X}, \mathbf{Y}, \mathbf{A}), \tag{1}$$

where p_θ models node features, p_γ the complete-data predictor, and p_λ the missingness mechanism. Conditioning on $\mathbf{A}$ makes explicit that the fixed graph structure can inform these distributions.

With incomplete node features, the predictive distribution is

$$p(\mathbf{Y} \mid \mathbf{X}_{\text{obs}}, \mathbf{M}, \mathbf{A}) = \int p(\mathbf{Y} \mid \mathbf{X}, \mathbf{M}, \mathbf{A})\, p(\mathbf{X}_{\text{mis}} \mid \mathbf{X}_{\text{obs}}, \mathbf{M}, \mathbf{A})\, d\mathbf{X}_{\text{mis}}. \tag{2}$$

A standard GNN parameterizes $p_\gamma(\mathbf{Y} \mid \mathbf{X}, \mathbf{A})$ for complete $\mathbf{X}$. It is, therefore, necessary to handle incomplete feature information before or during prediction.

Some methods handle missing node features without using topological information during feature recovery. They operate only on the incomplete feature matrix, typically by replacing missing entries in $\mathbf{X}_{\text{mis}}$ with statistics computed from the observed values of the same feature in $\mathbf{X}_{\text{obs}}$, such as the mean or median. For example, GCNmf models the feature distribution with a Gaussian mixture and uses the resulting conditional moments to handle missing coordinates (Taguchi et al., 2021). After filling missing entries with placeholders, GNN-MIM instead concatenates the missingness mask to the node features (Ferrini et al., 2026). In these approaches, the graph is still used by the downstream GNN, but missing features are not recovered using any topological information.

On the other hand, topology-aware approaches such as Feature Propagation (FP) (Rossi et al., 2022), estimates missing feature values by diffusing observed values across graph edges. For a feature channel j, let $\mathbf{x}^{(j)} = \mathbf{X}_{:,j}$, $O_j = \{i : M_{ij} = 0\}$, and $\mathcal{M}_j = \{i : M_{ij} = 1\}$ denote the feature vector and the nodes where feature j is observed and missing, respectively. FP repeatedly updates feature values through $\widehat{\mathbf{A}}$, the normalized adjacency matrix used for propagation, while restoring the observed entries after each step:

$$x_i^{(j,t+1)} = \begin{cases} X_{ij}, & i \in O_j, \\ \sum_{u \in V} \widehat{A}_{iu} x_u^{(j,t)}, & i \in \mathcal{M}_j. \end{cases} \tag{3}$$

PCFI extends this principle by assigning confidence scores based on the distance from observed features and combining channel-wise diffusion with inter-channel information (Um et al., 2023). FISF alternates propagation with repeated injection of the observed features (Um et al., 2025). Although these methods differ in their implementation, they all recover missing feature by propagating observed feature values through graph connections.

A crucial point often neglected in the literature is that the mechanisms used to describe the missingness are often orthogonal to how missingness is recovered. The latter is described by the missingness mechanism p_λ in equation 1. Following Rubin (Rubin, 1976), missingness mechanisms can be characterized through conditional dependencies between the mask and the data. Since $\mathbf{A}$ is fully observed and fixed in our setting, we express these conditions through $\mathbf{A}$.

For node classification, a mechanism is *feature-MAR* when

$$p(\mathbf{M} \mid \mathbf{X}_{\text{mis}}, \mathbf{X}_{\text{obs}}, \mathbf{A}) = p(\mathbf{M} \mid \mathbf{X}_{\text{obs}}, \mathbf{A}), \tag{4}$$

i.e., the missingness only depends on the observed features. On the other hand, is *label-MAR* when

$$p(\mathbf{M} \mid \mathbf{X}, \mathbf{Y}, \mathbf{A}) = p(\mathbf{M} \mid \mathbf{X}, \mathbf{A}), \tag{5}$$

and *Missing Completely At Random* (MCAR) when

$$p(\mathbf{M} \mid \mathbf{X}, \mathbf{Y}, \mathbf{A}) = p(\mathbf{M} \mid \mathbf{A}). \tag{6}$$

As in Ferrini et al. (2026) we treat the observed graph structure as fixed throughout learning and inference. Hence, MCAR refers to independence of the missingness mask from the random data variables X and Y, conditional on the realized graph A. Since $\mathbf{A}$ is fixed, equation 6 is equivalent to the usual statement that the mask is independent of the data $(\mathbf{X}, \mathbf{Y})$. Missingness is MNAR when the mask depends on hidden feature values or labels. Under feature-MAR and label-MAR, the missingness mechanism is ignorable for prediction and equation 2 reduces to

$$p(\mathbf{Y} \mid \mathbf{X}_{\text{obs}}, \mathbf{M}, \mathbf{A}) = \int p_\gamma(\mathbf{Y} \mid \mathbf{X}, \mathbf{A})\, p_\theta(\mathbf{X}_{\text{mis}} \mid \mathbf{X}_{\text{obs}}, \mathbf{A})\, d\mathbf{X}_{\text{mis}}. \tag{7}$$

To assess robustness to missing features, existing methods are commonly evaluated by injecting missingness into complete graph datasets. The standard protocol assumes uniform MCAR (U-MCAR), where each feature entry is removed independently with a fixed probability μ, $M_{ij} \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\mu)$. Another commonly used MCAR protocol removes the complete feature vectors of uniformly selected nodes; following Ferrini et al. (2026), we refer to this setting as Structural MCAR (S-MCAR).

The MCAR/MAR/MNAR taxonomy characterizes how $\mathbf{M}$ depends on the data, but not how the entries of $\mathbf{M}$ depend on one another. This distinction is crucial for graph-based recovery: two mechanisms can be equally MCAR and share the same global missingness rate, yet arrange missing entries according to the graph topology. This spatial arrangement directly affects propagation-based methods, as the configuration of observed entries determines the information available for diffusion. In the next section, we formalize this structural missingness when the graph shapes the mask.

3 GRAPH-STRUCTURED MISSINGNESS

To study how graph topology conditions missingness patterns, we first formalize the internal joint organization of the mask M. Let $p(\mathbf{M} \mid \mathbf{A})$ denote the joint distribution of the missingness mask conditional on the observed adjacency matrix $\mathbf{A}$. We say that the mask is *internally dependent* when its distribution does not factorize over individual entries:

$$p(\mathbf{M} \mid \mathbf{A}) \neq \prod_{i=1}^{n} \prod_{j=1}^{d} p(\mathbf{M}_{ij} \mid \mathbf{A}). \tag{8}$$

Equation 6 constrains the dependence between $\mathbf{M}$ and the data $(\mathbf{X}, \mathbf{Y})$, whereas Equation 8 concerns dependence among entries of $\mathbf{M}$: for example, consider an MCAR mechanism that masks exactly m out of the $N = nd$ feature entries uniformly without replacement. For any two distinct entries $a = ij$ and $b = zk$, $\Pr(\mathbf{M}_a = 1 \mid \mathbf{A}) = \frac{m}{N}$, while $\Pr(\mathbf{M}_b = 1 \mid \mathbf{M}_a = 1, \mathbf{A}) = \frac{m-1}{N-1} \neq \frac{m}{N}$. Thus, fixing the global missingness budget already induces dependence among mask entries and is a broad property. In this work, we are interested in a narrower case in which the graph itself shapes this dependence structure, which is a realistic and challenging scenario.

3.1 GRAPH-STRUCTURED DEPENDENCE WITHIN THE MISSINGNESS MASK

We characterize graph-structured dependence at the level of the joint mask distribution.

Definition 1 (Graph-structured mask dependence)**.** Let $p(\mathbf{M} \mid \mathbf{A})$ denote the conditional mask distribution. The missing mechanism exhibits *graph-structured mask dependence* if

1. the internal dependency on the mask is conditional on $\mathbf{A}$ and
2. the joint mask distribution depends on graph structure, i.e., there exist possible adjacency matrices $\mathbf{A}$ and $\mathbf{A}'$ on the same node set, each with positive probability, such that $p(\mathbf{M} \mid \mathbf{A}) \neq p(\mathbf{M} \mid \mathbf{A}')$, with all non-topological parameters of the mechanism held fixed.

This definition is broad enough to encompass the missingness mechanisms from the literature: U-MCAR without fixed budget is MCAR with internally independent mask; fixed-budget U-MCAR induces internally dependent masks without graph-structured dependence. Finally, S-MCAR also induces strong within-node dependence by masking complete feature vectors, but the masked nodes are sampled independently of $\mathbf{A}$. These standard protocols do not model graph-structured dependencies among mask entries. In the following, we introduce a mechanism designed to account for such dependencies.

3.2 RECURSIVE-TRANSMISSION MCAR

Graph-structured mask dependence can be instantiated in many ways. We propose a mechanism for graph-structured missingness called *Recursive Transmission* (RT) that: i) is directly comparable with standard MCAR protocols at the same missingness rate; ii) generates a missingness mask that can challenge existing propagation-based methods; iii) captures a plausible class of localized failure patterns. For a target missingness rate μ, let $m = \lfloor \mu nd \rfloor$ be the total number of entries to mask. We distribute this budget as evenly as possible across each feature channel j, obtaining budgets m_j such that $\sum_{j=1}^{d} m_j = m$. RT starts from a source node $s \in V$ shared across feature channels and expands each feature-specific missing region only through graph edges. For feature j with $m_j > 0$, initialize

$C_j^{(1)} = \{s\}$. At iteration t, define $\partial C_j^{(t)} = \left\{ v \notin C_j^{(t)} : \exists u \in C_j^{(t)} \text{ with } (u,v) \in E \right\}$, and, for each candidate $v \in \partial C_j^{(t)}$, $a_j^{(t)}(v) = \left| \mathcal{N}(v) \cap C_j^{(t)} \right|$.

The next node is chosen as $v_{t+1}^{(j)} \in \arg\max_{v \in \partial C_j^{(t)}} a_j^{(t)}(v)$, with uniform random tie-breaking, independently across feature channels. We then set $C_j^{(t+1)} = C_j^{(t)} \cup \{v_{t+1}^{(j)}\}$, and continue until $|C_j| = m_j$. Finally, $\mathbf{M}_{ij} = \mathbb{I}\{i \in C_j\}$. The structural differences between missingness protocols are illustrated in Figure 1. While U-MCAR removes feature entries uniformly at random and S-MCAR drops complete feature vectors for independently selected nodes, RT models localized failures that originate from a shared source node and expand along graph edges, generating compact, spatially correlated missing regions. The construction is motivated by the three goals above. First, RT represents a localized failure with a common origin, as may occur in sensor, communication, power-grid, or acquisition systems, while feature-specific growth allows different measurements to be affected along different trajectories.

Among the possible instantiations of the mechanism, we focus on the challenging case in which the seeds of the missingness of each feature coincide with the same source node. In general, in RT one could assign a distinct seed s to each feature. However, on large graphs, especially at moderate missingness rates and with relatively few features, this would place feature-specific missing regions in different parts of the graph. Many affected nodes would then lose only one or a few measurements, while retaining most of their feature vector. While such masks still contain graph-structured dependence, they describe a missingness that is easier to recover for methods that can exploit nearby observations or even just information from the other feature channels. On the other hand, by using the same seed for all channels, RT creates partially overlapping missing regions and therefore areas in which several measurements are simultaneously unavailable. Independent tie-breaking still allows the regions to diverge, avoiding to always end up in the more restrictive case where the same complete set of features is missing at every affected node.

Finally, using a single global rate μ makes the severity of RT directly comparable with fixed-budget U-MCAR and S-MCAR. The mechanisms can therefore be matched in the total number of missing entries while differing in how those entries are arranged over the graph. RT is not intended as a generative model of all real-world missingness patterns. Rather, it is a controlled stress test that isolates the effect of graph-localized mask dependence while matching standard protocols in the total missingness budget. In our experimental setting, we therefore complement RT with experiments on naturally occurring missingness, for which no graph-structured mechanism is imposed.

Proposition 1 (RT is graph-structured MCAR)**.** *If source selection and tie-breaking are independent of* $\mathbf{X}$ *and* $\mathbf{Y}$ *conditional on* $\mathbf{A}$*, then RT is MCAR and satisfies Definition 1.*

A formal proof is provided in Appendix A. RT is thus a realistic graph-structured MCAR mechanism whose particular geometry is chosen to probe a failure mode of propagation-based methods. We make this connection precise next.

3.3 Why Graph-Structured Missingness Is Difficult for Propagation

We now show why RT can be particularly challenging for methods that recover missing features through graph propagation. Consider one feature channel and omit the feature index for readability. Let O and M denote the observed and missing nodes for that feature, respectively. The FP update introduced in Equation 3 can be written in block form as

$$x_M^{(t+1)} = \widehat{\mathbf{A}}_{MM} x_M^{(t)} + \widehat{\mathbf{A}}_{MO} x_O. \tag{9}$$

The first term propagates the current estimates within the missing set, while the second introduces observed values into it. Define the observed boundary of the missing region as

$$\partial_O M = \{ u \in O : \exists v \in M \text{ with } (u,v) \in E \}.$$

Only nodes in $\partial_O M$ can directly inject observed information into the missing region.

Proposition 2 (Boundary bottleneck for feature propagation)**.** *Let* $\rho(\widehat{\mathbf{A}}_{MM})$ *denote the spectral radius of* $\widehat{\mathbf{A}}_{MM}$ *and assume* $\rho(\widehat{\mathbf{A}}_{MM}) < 1$*. Then the iteration in equation 9 converges to*

$$\widehat{x}_M = \mathbf{B} x_O, \qquad \mathbf{B} = \left(\mathbf{I} - \widehat{\mathbf{A}}_{MM} \right)^{-1} \widehat{\mathbf{A}}_{MO}. \tag{10}$$

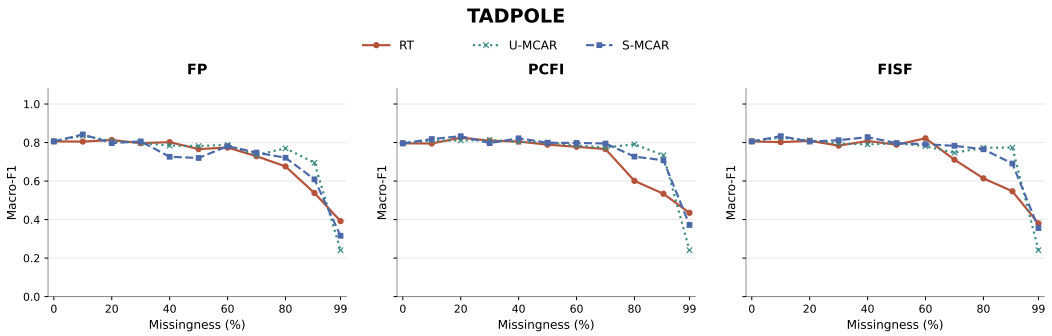


Figure 2: TADPOLE macro-F1 under RT, U-MCAR, and S-MCAR for FP, PCFI, and FISF. Curves show the mean and shaded bands indicate one standard deviation over five seeds. Complete macro-F1 and AUROC results are reported in Appendix C.

Moreover,

$$\operatorname{rank}(\mathbf{B}) \leq |\partial_O M|. \tag{11}$$

The proof is given in Appendix A. Equation 11 makes the relevant property of the mask explicit. When many missing nodes form a compact connected region with a small observed boundary, all reconstructed values inside the region are determined by only a limited number of observed inputs. In contrast, the same number of missing entries scattered across the graph may be adjacent to many more observed nodes.

To ground the theory of RT, we consider TADPOLE, which was identified in Ferrini et al. (2026) as a suitable benchmark for evaluating robustness to missing node features. We compare RT with U-MCAR and S-MCAR, all with the same global missingness rate. S-MCAR controls for strong within-node co-missingness while remaining independent of graph connectivity. Figure 2 shows that RT is more challenging for propagation-based methods. Averaging the macro-F1 AUC across FP, PCFI, and FISF gives 0.725 under RT, compared with 0.754 under U-MCAR and 0.751 under S-MCAR. Thus, graph-structured missingness reduces average propagation performance by approximately 4% relative to U-MCAR and 3.5% relative to S-MCAR.

4 PEMix: Structure-Conditioned Probabilistic Imputation

The analysis in Section 3.3 shows that propagation-based methods can become less effective when missingness propagates through graph connections. The graph, however, remains fully observed and can still provide information about node attributes. PEMix exploits this information without using topology to propagate observed feature values into missing regions.

Let $\mathbf{P} = g(\mathbf{A}) \in \mathbb{R}^{n \times q}$ be a fixed structural representation of the graph, with $\mathbf{p}_i$ denoting the representation of node i. In our experiments, g returns standardized non-trivial Laplacian eigenvectors. PEMix uses $\mathbf{P}$ to condition a probabilistic model of the incomplete node features, $p_\theta(\mathbf{X}^{\text{mis}} \mid \mathbf{X}^{\text{obs}}, \mathbf{A}) \approx q_\eta(\mathbf{X}^{\text{mis}} \mid \mathbf{X}^{\text{obs}}, \mathbf{P})$. The positional encodings are computed once from the fully observed graph and are not learned by the mixture.

Structure-Conditioned Mixture For each node i, let $\mathbf{x}_i = \mathbf{X}_{i,:} \in \mathbb{R}^d$ denote its feature vector and let $c_i \in \{1, \dots, K\}$ denote a latent mixture component, with $q(c_i = k) = \pi_k$. Conditioned on c_i, PEMix models node features and structural position independently with diagonal Gaussian distributions, $q_\eta(\mathbf{x}_i \mid c_i = k) = \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k^X, \operatorname{diag}((\boldsymbol{\sigma}_k^X)^2))$ and $q_\eta(\mathbf{p}_i \mid c_i = k) = \mathcal{N}(\mathbf{p}_i \mid \boldsymbol{\mu}_k^P, \operatorname{diag}((\boldsymbol{\sigma}_k^P)^2))$.

Let $\Omega_i = \{j \in \{1, \dots, d\} : M_{ij} = 0\}$ denote the observed coordinates of node i, and let $\bar{\Omega}_i = \{1, \dots, d\} \setminus \Omega_i$ denote the missing ones. Conditioning jointly on the observed coordinates and on structural position gives

$$\gamma_{ik} \propto \pi_k \, q_\eta(\mathbf{p}_i \mid c_i = k) \prod_{j \in \Omega_i} \mathcal{N}\left(X_{ij} \mid \mu_{kj}^X, (\sigma_{kj}^X)^2\right), \tag{12}$$

with normalization over k. Equivalently, the structural term induces a node-dependent prior $\pi_k(\mathbf{p}_i) = q_\eta(c_i = k \mid \mathbf{p}_i)$, which is then updated using the observed feature coordinates. The distinction from a feature-only mixture is therefore in the component responsibilities. Without structural conditioning, every node starts from the same prior π_k; in PEMix, structurally different nodes can have different component probabilities before their observed features are considered. This is particularly useful when only limited feature evidence remains. Under the diagonal covariance assumption, the conditional distribution of the missing coordinates is

$$q_\eta(\mathbf{x}_{i,\bar{\Omega}_i} \mid \mathbf{x}_{i,\Omega_i}, \mathbf{p}_i) = \sum_{k=1}^{K} \gamma_{ik} \prod_{j \in \bar{\Omega}_i} \mathcal{N}\left(X_{ij} \mid \mu_{kj}^X, (\sigma_{kj}^X)^2\right). \tag{13}$$

Thus, graph structure affects the distribution used to impute missing coordinates before message passing, rather than only determining how observed values are propagated between nodes.

End-to-End Prediction PEMix retains the conditional uncertainty in Equation 13 instead of committing to a single point estimate. Following Taguchi et al. (2021), the corresponding component-wise moments are marginalized analytically through the first message-passing layer, after which the remaining GNN layers are unchanged. The positional encodings affect only the mixture responsibilities in equation 12; they are not concatenated to the node representation passed to the downstream GNN. Following (Taguchi et al., 2021), who showed that jointly optimizing the mixture and downstream GNN can improve missing-feature reconstruction, we initialize the mixture parameters with EM and subsequently optimize them jointly with the GNN using the downstream task loss. We validate this design choice in Section F.3, where joint optimization consistently improves downstream performance compared with keeping the EM-initialized mixture fixed. Implementation details, including initialization, covariance parameterization, and positional-encoding dimension, are provided in Section E. Structural conditioning does not change the asymptotic cost of the probabilistic first layer: PEMix has the same per-epoch complexity as GCNmf, up to a dominated $O(nKq)$ term and a one-off partial eigendecomposition per graph (Section E.1).

5 Experimental Evaluation

We address three research questions:

- **RQ1:** Does PEMix improve robustness under challenging graph-structured missingness while remaining competitive under standard random missingness?
- **RQ2:** How does PEMix perform under naturally occurring missingness?
- **RQ3:** Are the gains of PEMix due simply to access to positional encodings, or specifically to using them to condition the missing-feature distribution?

Datasets and protocol. To answer RQ1, we consider four datasets with artificially injected missingness, namely TADPOLE (Zhu et al., 2019), GraphMSO, PEMS-BAY and Engrad (Marisca et al., 2024). Since these datasets are complete, the missingness mechanism and global missingness rate μ can be controlled directly. In particular we injected RT and U-MCAR with $\mu \in \{0, 0.1, \ldots, 0.9, 0.99\}$. For RQ2, we instead consider datasets where missingness is part of the original data. AQI, PV-US, and METR-LA are the air-quality, photovoltaic, and traffic datasets used by Marisca et al. (2024). We additionally evaluate PEMix in a different setting using RelBench (Robinson et al., 2024), a benchmark of relational databases with naturally incomplete features. For forecasting datasets, we convert each prediction time into an independent static graph sample using only past observations; the temporal windows, graph construction, and leakage-prevention protocol are detailed in Appendix B. The dataset overview is provided in Appendix B, Table 5. On the standard graph datasets, we compare against simple baseline with zero imputation (GNN-Zero), median imputation (GNN-Median) and mean imputation (GNN-MI), probabilistic and mask-aware approaches (GCNmf and GNN-MIM), and propagation methods (FP, PCFI, and FISF). RelBench requires a relational architecture and has substantially fewer missing-feature baselines. We therefore compare against the standard Relational Deep Learning model (RDL) (Robinson et al., 2024) and its missing-indicator extension RDL-MIM (Ferrini et al., 2026), and adapt PEMix to the same backbone, denoted RDL-PEMix. The datasets and evaluation regimes are summarized in Table 5.

Table 1: Performance under controlled RT and U-MCAR, reported as trapezoidal AUC over μ. Blue indicates the best model in each column, orange the second one, and green the third one. Full tables in Appendix D.

	RT				U-MCAR			
Model	TADPOLE Macro-F1 ↑	GraphMSO MAE ↓	PEMS-BAY MAE ↓	Engrad MAE ↓	TADPOLE Macro-F1 ↑	GraphMSO MAE ↓	PEMS-BAY MAE ↓	Engrad MAE ↓
GNN-Zero	0.7388	1.369627	4.294239	7.345723	0.7565	1.368397	4.059947	5.796827
GNN-Median	0.7015	1.369711	4.340867	7.317703	0.7398	1.368436	4.038923	5.859984
GNN-MI	0.7130	1.369657	4.251831	7.322302	0.7421	1.368595	4.079495	5.816402
FP	0.7254	1.371720	4.203665	6.515184	0.7473	1.370337	4.035542	**5.504311**
PCFI	0.7272	1.372185	4.308706	7.982851	0.7593	1.370070	4.115370	6.375060
FISF	0.7232	1.373884	4.237767	**6.468120**	0.7566	1.372575	4.071262	5.603992
GCNmf	0.7148	1.369023	3.921914	7.373967	0.7588	1.367092	**3.393461**	6.060386
GNN-MIM	0.7469	1.371282	4.193381	7.584628	0.7482	1.369593	3.909064	5.838703
PEMix	**0.8243**	**1.368784**	**3.764460**	7.206951	**0.8094**	**1.366935**	3.463835	5.753445

RQ1: PEMix gains most where graph-structured missingness is challenging, without sacrificing competitiveness under U-MCAR. For each controlled dataset, at each missingness rate, we evaluate all methods over five random seeds. The complete curves, including means and standard deviations at each rate, are reported in Section D. For compactness, in Table 1 we summarize each curve by its trapezoidal area under the curve (AUC) over the evaluated missingness rates. Higher AUC is better for macro-F1, while lower AUC is better for MAE. Under RT, the setting PEMix is explicitly designed to address, PEMix obtains the best AUC on three of the four datasets: TADPOLE, GraphMSO, and PEMS-BAY, while ranking third on Engrad. The gains are particularly pronounced on TADPOLE and PEMS-BAY, where PEMix improves substantially over the strongest competing baselines. This supports the intended role of structural conditioning when missing observations form graph-localized regions and local propagation becomes less informative. Under U-MCAR, in contrast, we would not necessarily expect PEMix to dominate, since missing entries are scattered independently and propagation-based or simpler recovery mechanisms are not exposed to the same structural bottleneck. Nevertheless, PEMix remains competitive: it ranks first on TADPOLE and GraphMSO, second on PEMS-BAY, and third on EngRAD. Thus, the benefit of PEMix is concentrated in the challenging graph-structured regime for which it was designed, while its use does not generally entail a loss of competitiveness when missingness is unstructured.

RQ2: PEMix performs strongly under naturally occurring missingness.

Results on datasets with naturally occurring missingness are reported in Table 2. PEMix achieves the lowest MAE on two of the three benchmarks, AQI and METR-LA, while ranking second on PV-US. This pattern is consistent with the spatial organization of missingness in these datasets. We quantify this structure through the excess co-missingness of neighboring nodes: intuitively, we compare how often the two endpoints of an edge are simultaneously missing with how often this would occur under spatially independent missingness. Values above one therefore indicate that missing observations cluster over the graph. The exact definition and dataset-level values are reported in Appendix B.1. All three datasets exhibit co-missingness above the range expected under spatially independent missingness. The effect is stronger on METR-LA, whose ratio reaches 4.03 over timestamos with partial missingness, compared with 1.16 on AQI and 1.21 on PV-US. PEMix achieves the best MAE on AQI and METR-LA, whereas on PV-US GCNmf performs best and PEMix remains a close second. The same trend extends beyond standard graph benchmarks: in the relational setting, RDL-PEMix obtains the best result on four of the six RelBench tasks reported in Table 3, suggesting that PEMix can be integrated into relational models as an effective strategy for handling missing features.

Table 2: Performance with naturally occurring missingness (test MAE, mean $\pm$ std over five seeds; lower is better). Blue, orange, and green mark the three best models within each dataset.

Model	AQI	PV-US	METR-LA
GNN-Zero	20.1895 ± 0.2203	12.4871 ± 0.0151	6.5199 ± 0.0465
GNN-Median	19.7317 ± 0.2595	12.3410 ± 0.0289	6.6117 ± 0.0760
GNN-MI	19.9730 ± 0.1712	12.4844 ± 0.0337	6.5197 ± 0.0576
FP	18.9450 ± 0.0409	12.5251 ± 0.0290	6.5945 ± 0.0487
PCFI	19.5384 ± 0.0707	12.9281 ± 0.0698	6.6484 ± 0.0957
FISF	19.0586 ± 0.0778	12.5746 ± 0.0203	6.5991 ± 0.0721
GCNmf	18.9071 ± 0.0460	$\mathbf{11.9906 \pm 0.1444}$	6.5170 ± 0.0182
GNN-MIM	19.4978 ± 0.1793	12.2302 ± 0.0135	6.7227 ± 0.0778
PEMix	$\mathbf{18.7274 \pm 0.0523}$	12.0225 ± 0.0392	$\mathbf{6.4511 \pm 0.0105}$

Table 3: RelBench results with naturally occurring missingness (mean ± std over three seeds; ten epochs). RDL-PEMix denotes the relational adaptation of PEMix, while RDL-MIM is the missing-indicator adaptation of RDL. Bold marks the best mean within each task.

Dataset	Task	Metric	RDL	RDL-MIM	RDL-PEMix
Node regression					
REL-TRIAL	study-adverse	MAE↓	44.470 ± 0.210	43.115 ± 0.155	$\mathbf{42.840 \pm 0.680}$
REL-TRIAL	site-success	MAE↓	0.400 ± 0.020	$\mathbf{0.375 \pm 0.017}$	0.391 ± 0.008
REL-F1	driver-position	MAE↓	$\mathbf{4.022 \pm 0.120}$	4.0271 ± 0.0389	4.208 ± 0.090
Node classification					
REL-TRIAL	study-outcome	AUC↑	0.686 ± 0.010	0.6794 ± 0.0051	$\mathbf{0.706 \pm 0.010}$
REL-F1	driver-top3	AUC↑	0.755 ± 0.006	0.768 ± 0.004	$\mathbf{0.773 \pm 0.010}$
REL-STACK	user-engagement	AUC↑	0.9062 ± 0.0002	0.9060 ± 0.0008	$\mathbf{0.9072 \pm 0.0001}$

RQ3: Structural conditioning, rather than access to PE alone, drives PEMix's gains.

To isolate the role of positional encodings, we compare two alternatives. First, we provide the same PE to propagation-based methods inside their missing-feature mechanism. Second, we concatenate PE only after missing-feature handling, before the downstream predictor. We also include a PE-only control. We conducted this ablation on the Tadpole and PV-US dataset. As in RQ1, for the Tadpole dataset we summarize each missingness curve by its macro-F1 AUC; complete results with average and standard deviations over five seeds for each level of missingness μ are reported in Appendix F. For PV-US we report means and standard deviations over five seeds. Results are reported in Table 4. Simply providing positional encodings is not sufficient to reproduce PEMix's performance. Adding the same PE to propagation-based missing-feature mechanisms leaves a substantial gap, while using PE only downstream improves performance on TADPOLE but remains below PEMix. The PE-only control performs poorly on both datasets. These results suggest that PEMix benefits specifically from using structural information to condition the missing-feature distribution, rather than from access to positional encodings per se.

Table 4: Effect of PE on TADPOLE and PV-US. "Mechanism" denotes methods in which PE are used while handling missing features, whereas "Downstream" denotes variants where PE are concatenated only after missing-feature handling. We report macro-F1 AUC across missingness rates for TADPOLE (higher is better) and test MAE under natural missingness for PV-US (lower is better). PV-US results are means ± std over five seeds.

PE usage	Model	TADPOLE Macro-F1 AUC ↑	PV-US MAE ↓
Mechanism	PEMix	**0.8243**	$\mathbf{12.0225 \pm 0.0392}$
	FP + PE	0.7336	12.4642 ± 0.0221
	PCFI + PE	0.7474	12.8643 ± 0.0454
	FISF + PE	0.7573	12.5981 ± 0.0443
Downstream	GCNmf + PE	0.7211	12.1421 ± 0.0145
	FP + PE	0.7754	12.1261 ± 0.0452
	PCFI + PE	0.7745	12.6979 ± 0.0790
	FISF + PE	0.7798	12.2364 ± 0.0649
–	PE only	0.2923	16.5526 ± 0.0401

6 Conclusion

We studied missing node features when missingness itself is structured by the graph. We formalized graph-structured mask dependence and introduced RT, a controlled mechanism that exposes a boundary bottleneck of propagation-based method. Motivated by this limitation, we proposed PEMix, which conditions a probabilistic model of incomplete node features on global structural representations. Across controlled and naturally occurring missingness, PEMix is particularly effective when missing observations are graph-localized, while remaining competitive under standard missingness. Ablations further show that the gains arise from using structural information to condition the missing-feature distribution, rather than from access to positional encodings alone. These results highlight graph structure as a useful source of information not only for prediction, but also for modeling incomplete node features. **Limitations:** PEMix assumes diagonal mixture covariances and a fully observed graph, and our evaluation is limited to node classification and forecasting. Broader analyses of structural encodings, temporal modeling, and other downstream tasks are left for future work.

AI USE STATEMENT

In this work, we used OpenAI Codex with the GPT-5.6 Sol model to draft portions of the manuscript, edit the text for clarity and readability, and assist in proving mathematical claims, including suggesting proof strategies and drafting portions of the proofs presented in Appendix A. We did not use generative AI to formulate the mathematical claims themselves. The authors independently checked every step of each AI-assisted proof and revised the final arguments. All AI-assisted text was reviewed and edited by the authors.

We did not use generative AI for any other task requiring disclosure under the ICLR 2027 AI Policy, including developing theoretical models or conceptual frameworks, proposing or refining hypotheses, designing or providing feedback on the research methodology or experiments, implementing methods, generating synthetic data, cleaning or reformatting data, translation, qualitative or thematic data analysis, or interpreting results. Generative AI was not used to run experiments or produce the reported experimental results. We take responsibility for the final content of this work, including all text, mathematical claims, proofs, and other artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

An anonymized implementation, configuration files, and scripts for reproducing the reported experiments are available at `https://anonymous.4open.science/r/4CE7/`. Section 2 defines the missing-feature learning problem and the missingness taxonomy. Definition 1 and Section 3.2 define graph-structured mask dependence and the RT mechanism. Section 3.3 establishes the boundary bottleneck for feature propagation, while Section 4 defines PEMix and its structure-conditioned probabilistic model. Appendix A provides the proofs and technical derivations. Appendix B reports the datasets, preprocessing procedures, missingness protocols, evaluation metrics, data splits, and random seeds. Appendix E documents the optimization settings, covariance assumptions, hyperparameters, and computational-complexity analysis. Appendices C and F provide the additional propagation controls and ablation studies.

REFERENCES

Carlijn I. R. Braem, Utku S. Yavuz, Hermie J. Hermens, and Peter H. Veltink. Missing data statistics provide causal insights into data loss in diabetes health monitoring by wearable sensors. *Sensors*, 24 (5):1526, 2024. doi: 10.3390/S24051526. URL `https://doi.org/10.3390/s24051526`.

Rahmat Nur Faizin, Mardhani Riasetiawan, and Ahmad Ashari. A review of missing sensor data imputation methods. In *2019 5th International Conference on Science and Technology (ICST)*, volume 1, pp. 1–6, 2019. doi: 10.1109/ICST47872.2019.9166287.

Francesco Ferrini, Veronica Lachi, Antonio Longa, Bruno Lepri, Akiyoshi Matono, Andrea Passerini, Xin Liu, and Manfred Jaeger. Rethinking GNNs and missing features: Challenges, evaluation and a robust solution. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=E4gftzjqoh`.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 1024–1034, 2017. URL `https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html`.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL `https://openreview.net/forum?id=SJU4ayYgl`.

Ivan Marisca, Cesare Alippi, and Filippo Maria Bianchi. Graph-based forecasting with missing data through spatiotemporal downsampling. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27,*

2024, volume 235 of *Proceedings of Machine Learning Research*, pp. 34846–34865. PMLR / OpenReview.net, 2024. URL https://proceedings.mlr.press/v235/marisca24a.html.

Joshua Robinson, Rishabh Ranjan, Weihua Hu, Kexin Huang, Jiaqi Han, Alejandro Dobles, Matthias Fey, Jan Eric Lenssen, Yiwen Yuan, Zecheng Zhang, Xinwei He, and Jure Leskovec. Relbench: A benchmark for deep learning on relational databases. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/25cd345233c65fac1fec0ce61d0f7836-Abstract-Datasets_and_Benchmarks_Track.html.

Emanuele Rossi, Henry Kenlay, Maria I. Gorinova, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. On the unreasonable effectiveness of feature propagation in learning on graphs with missing node features. In Bastian Rieck and Razvan Pascanu (eds.), *Learning on Graphs Conference, LoG 2022, 9-12 December 2022, Virtual Event*, volume 198 of *Proceedings of Machine Learning Research*, pp. 11. PMLR, 2022. URL https://proceedings.mlr.press/v198/rossi22a.html.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Hibiki Taguchi, Xin Liu, and Tsuyoshi Murata. Graph convolutional networks for graphs containing missing features. *Future Gener. Comput. Syst.*, 117:155–168, 2021. doi: 10.1016/J.FUTURE.2020.11.016. URL https://doi.org/10.1016/j.future.2020.11.016.

Daeho Um, Jiwoong Park, Seulki Park, and Jin Young Choi. Confidence-based feature imputation for graphs with partially known features. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=YPKBIILy-Kt.

Daeho Um, Sunoh Kim, Jiwoong Park, Jongin Lim, Seong-Jin Ahn, and Seulki Park. Propagate and inject: Revisiting propagation-based feature imputation for graphs with partially observed features. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025. URL https://proceedings.mlr.press/v267/um25a.html.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Qikui Zhu, Bo Du, and Pingkun Yan. Multi-hop convolutions on weighted graphs. *CoRR*, abs/1911.04978, 2019. URL http://arxiv.org/abs/1911.04978.

A PROOFS

A.1 RT IS GRAPH-STRUCTURED MCAR

Proposition 1 (RT is graph-structured MCAR), restated. If source selection and tie-breaking are independent of $\mathbf{X}$ and $\mathbf{Y}$ conditional on $\mathbf{A}$, then RT is MCAR in the sense of equation 6. Under the non-degeneracy and graph-support conditions stated in Appendix A, it also satisfies Definition 1.

Proof. We work conditional on the realized adjacency matrix $\mathbf{A}$ and fixed feature budgets, as in the main text. For the graph-structured conclusion, we make the necessary non-degeneracy conditions explicit: the graph is connected and undirected, the source distribution has full support on V, and some channel has $2 \le m_j \le n-1$. For the topology comparison below, the complete graph and the star on V are assumed to have positive probability under the distribution of $\mathbf{A}$. In comparing graphs, the source law and all budgets are held fixed. These conditions are needed for that conclusion: at zero or full missingness the mask is deterministic, and if all budgets are at most one, no growth step need consult an edge. Thus independence from the data alone proves MCAR, while the graph-structured assertion is understood in this non-degenerate regime.

MCAR. Let U collect the source and all tie-breaking variables. The recursion is a deterministic map $\mathbf{M} = F(\mathbf{A}, U)$: each update uses only adjacency, the current region, and U. Since U is independent of $(\mathbf{X}, \mathbf{Y})$ conditional on $\mathbf{A}$, for every mask m,

$$\Pr(\mathbf{M} = m \mid \mathbf{X}, \mathbf{Y}, \mathbf{A}) = \Pr(F(\mathbf{A}, U) = m \mid \mathbf{A}) = \Pr(\mathbf{M} = m \mid \mathbf{A}).$$

This is Equation 6; dependence within the mask does not affect this conditional independence from the data.

Internal dependence. Fix a channel with $2 \le m_j \le n-1$. Every node has positive probability of being masked, since it can be the source. The fixed budget implies $\sum_i M_{ij} = m_j < n$ almost surely, so at least one indicator also has positive probability of being zero. If the channel indicators were mutually independent, then

$$0 = \mathrm{Var}\left(\sum_i M_{ij} \mid \mathbf{A}\right) = \sum_i \mathrm{Var}(M_{ij} \mid \mathbf{A}) > 0,$$

which is impossible. Independence of every entry of the full mask would imply independence within this channel. The conditional distribution $p(\mathbf{M} \mid \mathbf{A})$ therefore does not factorize.

Dependence on topology. Compare the complete graph with adjacency matrix $\mathbf{A}$ on the labeled set V with a star centered at a with adjacency matrix $\mathbf{A}'$, holding the full-support source law, budgets, and uniform tie-breaking rule fixed. Choose $S \subseteq V \setminus \{a\}$ with $|S| = m_j$. In the complete graph all boundary candidates have the same score. Starting in S and then selecting its remaining nodes has positive probability, so $\Pr(C_j = S \mid \mathbf{A}) > 0$. On the star every RT region is connected, because it starts with one node and grows by adding boundary nodes. A set of at least two leaves excluding the center is disconnected, so $\Pr(C_j = S \mid \mathbf{A}') = 0$. The conditional channel marginals, and hence the joint mask distributions $p(\mathbf{M} \mid \mathbf{A})$ and $p(\mathbf{M} \mid \mathbf{A}')$, differ. Together with internal dependence, this proves Definition 1 under the stated non-degeneracy conditions. □

A.2 BOUNDARY BOTTLENECK FOR FEATURE PROPAGATION

Proposition 2 (Boundary bottleneck for feature propagation), restated. Let $\rho(\widehat{\mathbf{A}}_{MM})$ denote the spectral radius of $\widehat{\mathbf{A}}_{MM}$ and assume $\rho(\widehat{\mathbf{A}}_{MM}) < 1$. Then the iteration in equation 9 converges to

$$\widehat{x}_M = \mathbf{B} x_O, \qquad \mathbf{B} = \left(\mathbf{I} - \widehat{\mathbf{A}}_{MM}\right)^{-1} \widehat{\mathbf{A}}_{MO}.$$

Moreover,

$$\mathrm{rank}(\mathbf{B}) \le |\partial_O M|.$$

Proof. Conditional on the realized graph, the normalized adjacency is a fixed local propagation matrix: an off-diagonal entry can be nonzero only on a graph edge. Observations x_O remain clamped.

Set $T = \widehat{\mathbf{A}}_{MM}$ and $C = \widehat{\mathbf{A}}_{MO}$. For every integer $t \geq 1$, induction on the update gives

$$x_M^{(t)} = T^t x_M^{(0)} + \sum_{r=0}^{t-1} T^r C x_O.$$

Since $\rho(T) < 1$, every Jordan block of T^t tends to zero: its entries are finite sums of polynomial factors in t multiplied by powers of eigenvalues of modulus less than one (nilpotent blocks vanish after finitely many powers). Consequently $T^t \to 0$. Also $I - T$ is invertible, and the finite geometric-series identity gives

$$\sum_{r=0}^{t-1} T^r = (I - T)^{-1}(I - T^t) \longrightarrow (I - T)^{-1}.$$

For any finite initial vector, the initial-condition term vanishes and the limit is $(I-T)^{-1}Cx_O = \mathbf{B}x_O$. Invertibility also makes this the unique fixed point.

If $u \in O \setminus \partial_O M$, there is no edge from a missing node to u, so locality gives $C_{:,u} = 0$. Left multiplication by $(I - T)^{-1}$ preserves that zero column. Thus only observed boundary coordinates can influence the limit. Since C has at most $|\partial_O M|$ nonzero columns,

$$\operatorname{rank}(\mathbf{B}) = \operatorname{rank}(C) \leq |\partial_O M|,$$

where the equality follows from invertibility of $I - T$. □

The result concerns fixed linear, channel-wise propagation. It does not automatically extend to nonlinear or cross-channel recovery, and a low-rank recovery map alone does not imply positive error for every graph signal.

B Datasets and Evaluation Protocol

Static formulation of forecasting datasets. For AQI, PV-US, METR-LA, and GraphMSO, we convert each prediction time into an independent static graph sample. The graph topology is fixed over time and constructed from sensor locations, official distance files, or the GraphMSO generator, depending on the dataset; it never uses future observations. Node features contain only past sensor measurements, flattened over the input window, together with unmasked calendar covariates when available. Targets are the subsequent measurements. The input/output windows are 24/6 for AQI, 72/6 for PV-US, 24/12 for METR-LA, and 72/36 for GraphMSO. No temporal model is used: all methods operate on the resulting static graph samples with the same inputs. To prevent leakage, train/validation/test splits are chronological and windows are constructed so that each input–target pair lies entirely within one split. Standardization and imputation statistics are estimated from observed training values only. Artificial masks are generated on the raw series before windowing, so repeated occurrences of the same measurement share the same missingness state. Naturally missing values are preserved, and missing targets are excluded from both loss and evaluation. PEMix positional encodings depend only on the fixed graph and are computed once per dataset; propagation-based imputers operate only on sensor measurements, not on calendar covariates.

Table 5: Datasets used in the controlled and native-missingness experiments. Controlled datasets are originally complete and missingness is injected artificially. Native datasets contain missing values in the original data. For native datasets, we report the percentage of missing entries in the original data.

Regime	Dataset	Domain	Source	Missing (%)
Controlled	TADPOLE	clinical	Zhu et al. (2019)	–
	GraphMSO	graph forecasting	Marisca et al. (2024)	–
	PEMS-BAY	traffic forecasting	Marisca et al. (2024)	–
	Engrad	forecasting	Marisca et al. (2024)	–
Native	AQI	air-quality forecasting	Marisca et al. (2024)	25.67
	PV-US	photovoltaic forecasting	Marisca et al. (2024)	54.02
	METR-LA	traffic forecasting	Marisca et al. (2024)	8.11
	RelBench (rel-f1)	relational databases	Robinson et al. (2024)	9.42
	RelBench (rel-trial)	relational databases	Robinson et al. (2024)	23.90
	RelBench (rel-stack)	relational databases	Robinson et al. (2024)	19.13

Table 6: Co-missingness ratio HC_{miss} on datasets with naturally occurring missingness. Values above 1 indicate that neighboring nodes are missing together more often than under spatially independent missingness. The permutation null is the central 95% range over 999 node permutations, computed over all timestamps.

Dataset	All timestamps	Partial timestamps	Permutation null (95%)
AQI	1.15	1.16	[0.902, 1.118]
PV-US	1.03	1.21	[0.999, 1.001]
METR-LA	1.04	4.03	[0.997, 1.003]

The controlled experiments comprise one classification dataset, namely TADPOLE, and three forecasting datasets (GraphMSO, PEMS-BAY, and Engrad). Native missingness is evaluated on AQI, PV-US, METR-LA, and the six RelBench tasks in Table 3.

Controlled curves are integrated with the trapezoidal rule over $\mu \in \{0, 0.1, \ldots, 0.9, 0.99\}$, without division by 0.99. Classification results use five seeds (1, 43, 15, 118, 222), with the same splits for every model and missingness mechanism; forecasting uses five test seeds per point. Native forecasting uses five seeds, and RelBench uses three seeds and ten epochs. AUC differences are reported descriptively; we do not perform significance tests.

B.1 Spatial organization of missingness in real-world datasets

We test whether naturally occurring missing values fall on neighboring nodes more often than they would if missingness were spatially independent.

Co-missingness ratio. Let $M_{tij} = 1$ if feature j of node i is missing at timestamp t. For each feature j and timestamp t, let

$$p_{tj} = \frac{1}{n}\sum_{i=1}^{n} M_{tij}, \qquad h_{tj}^{\text{miss}} = \frac{1}{|E|}\sum_{(u,v)\in E} \mathbb{I}[M_{tuj} = 1 \wedge M_{tvj} = 1]$$

denote the missingness rate and the fraction of edges whose two endpoints are both missing. If missingness were independent across nodes with the same rate p_{tj}, the expected fraction of missing–missing edges would be p_{tj}^2. Over a set of timestamps $\mathcal{T}$, we therefore define

$$HC_{\text{miss}}(j) = \frac{\sum_{t\in\mathcal{T}} h_{tj}^{\text{miss}}}{\sum_{t\in\mathcal{T}} p_{tj}^2},$$

a ratio of sums rather than an average of per-timestamp ratios. $HC_{\text{miss}} = 1$ corresponds to spatial independence, and values above 1 indicate excess co-missingness between neighboring nodes. Each dataset has a single measured channel (PM2.5 for AQI, power for PV-US, and speed for METR-LA), so the dataset-level value is that of its channel.

Evaluation subsets. We compute the ratio over all timestamps and over *partial* timestamps, those with $0 < p_{tj} < 1$. Completely observed and completely missing snapshots carry no spatial contrast. This distinction matters on METR-LA, where many missing values come from graph-wide outages.

Permutation control. To check that excess co-missingness reflects graph position rather than marginal missingness rates, we randomly permute the complete missingness trajectories of nodes across graph positions. This preserves temporal missingness patterns and the number of missing nodes at each timestamp, while breaking their association with the topology. We use 999 permutations over all timestamps and report the central 95% range of the resulting null distribution.

Results. Table 6 reports the results. On all three datasets, the observed ratio lies above its permutation null range, so co-missingness is associated with graph position. Over all timestamps the excess is largest on AQI; over partial timestamps it is largest on METR-LA (4.03), followed by PV-US (1.21) and AQI (1.16).

C PROPAGATION CONTROLS ON TADPOLE

Tables 7–9 give the complete macro-F1 and classification AUROC results for Figure 2 (mean $\pm$ standard deviation over five seeds).

For a metric $s(\mu)$, we integrate its mean curve as $\sum_k (\mu_{k+1} - \mu_k)[s(\mu_k) + s(\mu_{k+1})]/2$ over $[0, 0.99]$, without normalisation. Integrating the rounded table means gives average macro-F1 areas of 0.7252, 0.7543, and 0.7508 under RT, U-MCAR, and S-MCAR, respectively. The corresponding areas under the classification-AUROC curves are 0.8855, 0.8958, and 0.8963. These curve areas are distinct from the classification AUROC at a particular missingness rate.

D COMPLETE RESULTS UNDER CONTROLLED MISSINGNESS

The following tables report macro-F1 for TADPOLE and MAE for GraphMSO, PEMS-BAY, and EngRAD. Each dataset is split across two tables (0–50% and 60–99%) for legibility. hading marks the three best trapezoidal AUC values within each dataset.

Table 7: Complete FP results on TADPOLE (five seeds). Higher macro-F1 and AUROC are better.

	Macro-F1			AUROC		
μ (%)	RT	U-MCAR	S-MCAR	RT	U-MCAR	S-MCAR
0	0.806 ± 0.065	0.806 ± 0.065	0.806 ± 0.065	0.929 ± 0.022	0.929 ± 0.022	0.929 ± 0.022
10	0.805 ± 0.022	0.836 ± 0.038	0.842 ± 0.028	0.932 ± 0.013	0.939 ± 0.016	0.939 ± 0.019
20	0.813 ± 0.035	0.808 ± 0.044	0.797 ± 0.038	0.941 ± 0.010	0.934 ± 0.021	0.932 ± 0.014
30	0.797 ± 0.034	0.798 ± 0.037	0.806 ± 0.065	0.935 ± 0.015	0.928 ± 0.016	0.924 ± 0.031
40	0.802 ± 0.054	0.784 ± 0.026	0.726 ± 0.086	0.926 ± 0.019	0.924 ± 0.013	0.897 ± 0.022
50	0.766 ± 0.060	0.782 ± 0.051	0.720 ± 0.116	0.917 ± 0.019	0.925 ± 0.027	0.906 ± 0.026
60	0.774 ± 0.075	0.788 ± 0.044	0.781 ± 0.040	0.931 ± 0.018	0.921 ± 0.031	0.918 ± 0.018
70	0.729 ± 0.144	0.735 ± 0.048	0.748 ± 0.064	0.893 ± 0.091	0.897 ± 0.010	0.908 ± 0.014
80	0.676 ± 0.120	0.770 ± 0.052	0.721 ± 0.066	0.862 ± 0.051	0.915 ± 0.028	0.890 ± 0.028
90	0.538 ± 0.078	0.695 ± 0.060	0.609 ± 0.118	0.811 ± 0.068	0.904 ± 0.030	0.872 ± 0.049
99	0.392 ± 0.074	0.241 ± 0.000	0.316 ± 0.103	0.673 ± 0.044	0.492 ± 0.027	0.604 ± 0.119

Table 8: Complete PCFI results on TADPOLE (five seeds). Higher macro-F1 and AUROC are better.

	Macro-F1			AUROC		
μ (%)	RT	U-MCAR	S-MCAR	RT	U-MCAR	S-MCAR
0	0.796 ± 0.054	0.796 ± 0.054	0.796 ± 0.054	0.927 ± 0.016	0.927 ± 0.016	0.927 ± 0.016
10	0.795 ± 0.020	0.810 ± 0.046	0.818 ± 0.038	0.928 ± 0.008	0.936 ± 0.023	0.938 ± 0.015
20	0.825 ± 0.032	0.811 ± 0.038	0.833 ± 0.021	0.940 ± 0.018	0.938 ± 0.020	0.940 ± 0.013
30	0.810 ± 0.019	0.813 ± 0.043	0.797 ± 0.036	0.938 ± 0.017	0.934 ± 0.021	0.928 ± 0.022
40	0.805 ± 0.047	0.804 ± 0.028	0.822 ± 0.047	0.929 ± 0.016	0.933 ± 0.011	0.929 ± 0.016
50	0.789 ± 0.059	0.800 ± 0.035	0.799 ± 0.023	0.922 ± 0.021	0.927 ± 0.020	0.931 ± 0.017
60	0.778 ± 0.078	0.785 ± 0.049	0.798 ± 0.016	0.924 ± 0.025	0.922 ± 0.024	0.923 ± 0.009
70	0.766 ± 0.093	0.774 ± 0.036	0.795 ± 0.039	0.910 ± 0.046	0.917 ± 0.011	0.921 ± 0.013
80	0.601 ± 0.166	0.791 ± 0.050	0.727 ± 0.117	0.828 ± 0.079	0.926 ± 0.024	0.899 ± 0.038
90	0.534 ± 0.173	0.734 ± 0.046	0.708 ± 0.113	0.762 ± 0.207	0.911 ± 0.028	0.912 ± 0.018
99	0.435 ± 0.067	0.241 ± 0.000	0.372 ± 0.125	0.725 ± 0.015	0.487 ± 0.105	0.626 ± 0.175

Table 9: Complete FISF results on TADPOLE (five seeds). Higher macro-F1 and AUROC are better.

	Macro-F1			AUROC		
μ (%)	RT	U-MCAR	S-MCAR	RT	U-MCAR	S-MCAR
0	0.806 ± 0.040	0.806 ± 0.040	0.806 ± 0.040	0.930 ± 0.016	0.930 ± 0.016	0.930 ± 0.016
10	0.802 ± 0.013	0.825 ± 0.058	0.833 ± 0.021	0.932 ± 0.015	0.939 ± 0.019	0.940 ± 0.014
20	0.809 ± 0.041	0.811 ± 0.036	0.805 ± 0.042	0.940 ± 0.009	0.932 ± 0.017	0.934 ± 0.012
30	0.784 ± 0.055	0.795 ± 0.045	0.812 ± 0.069	0.930 ± 0.018	0.930 ± 0.016	0.924 ± 0.028
40	0.808 ± 0.049	0.790 ± 0.045	0.828 ± 0.032	0.932 ± 0.013	0.933 ± 0.014	0.928 ± 0.010
50	0.789 ± 0.040	0.797 ± 0.035	0.797 ± 0.037	0.924 ± 0.011	0.934 ± 0.022	0.919 ± 0.027
60	0.822 ± 0.044	0.781 ± 0.060	0.791 ± 0.042	0.935 ± 0.014	0.915 ± 0.023	0.927 ± 0.012
70	0.711 ± 0.146	0.747 ± 0.079	0.783 ± 0.045	0.905 ± 0.047	0.907 ± 0.022	0.920 ± 0.023
80	0.614 ± 0.145	0.773 ± 0.054	0.765 ± 0.045	0.834 ± 0.088	0.925 ± 0.021	0.917 ± 0.010
90	0.547 ± 0.145	0.774 ± 0.061	0.691 ± 0.098	0.801 ± 0.062	0.917 ± 0.027	0.887 ± 0.044
99	0.380 ± 0.137	0.241 ± 0.000	0.356 ± 0.158	0.665 ± 0.098	0.476 ± 0.066	0.607 ± 0.213

Table 10: TADPOLE: complete macro-F1 results under RT (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation.

Model	0	10	20	30	40	50
GNN-Zero	0.825 ± 0.038	0.811 ± 0.026	0.807 ± 0.052	0.804 ± 0.039	0.815 ± 0.034	0.761 ± 0.066
GNN-Median	0.825 ± 0.038	0.797 ± 0.030	0.800 ± 0.038	0.789 ± 0.070	0.802 ± 0.045	0.782 ± 0.055
GNN-MI	0.825 ± 0.038	0.805 ± 0.033	0.811 ± 0.044	0.806 ± 0.033	0.812 ± 0.051	0.777 ± 0.035
FP	0.806 ± 0.065	0.805 ± 0.022	0.813 ± 0.035	0.797 ± 0.034	0.802 ± 0.054	0.766 ± 0.060
PCFI	0.796 ± 0.054	0.795 ± 0.020	0.825 ± 0.032	0.810 ± 0.019	0.805 ± 0.047	0.789 ± 0.059
FISF	0.806 ± 0.040	0.802 ± 0.013	0.809 ± 0.041	0.784 ± 0.055	0.808 ± 0.049	0.789 ± 0.040
GCNmf	0.814 ± 0.043	0.807 ± 0.051	0.818 ± 0.071	0.796 ± 0.070	0.836 ± 0.056	0.770 ± 0.154
GNN-MIM	0.828 ± 0.038	0.801 ± 0.052	0.820 ± 0.025	0.797 ± 0.054	0.824 ± 0.050	0.798 ± 0.057
PEMix	0.828 ± 0.042	0.864 ± 0.041	0.846 ± 0.050	0.838 ± 0.050	0.844 ± 0.054	0.826 ± 0.028

Table 11: TADPOLE: complete macro-F1 results under RT (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (higher is better).

Model	60	70	80	90	99	AUC
GNN-Zero	0.808 ± 0.049	0.788 ± 0.083	0.680 ± 0.100	0.534 ± 0.079	0.429 ± 0.049	0.7388
GNN-Median	0.764 ± 0.063	0.714 ± 0.138	0.589 ± 0.165	0.470 ± 0.076	0.267 ± 0.057	0.7015
GNN-MI	0.785 ± 0.041	0.749 ± 0.101	0.616 ± 0.113	0.473 ± 0.075	0.241 ± 0.000	0.7130
FP	0.774 ± 0.075	0.729 ± 0.144	0.676 ± 0.120	0.538 ± 0.078	0.392 ± 0.074	0.7254
PCFI	0.778 ± 0.078	0.766 ± 0.093	0.601 ± 0.166	0.534 ± 0.173	0.435 ± 0.067	0.7272
FISF	0.822 ± 0.044	0.711 ± 0.146	0.614 ± 0.145	0.547 ± 0.145	0.380 ± 0.137	0.7232
GCNmf	0.776 ± 0.148	0.721 ± 0.151	0.601 ± 0.149	0.482 ± 0.022	0.354 ± 0.064	0.7148
GNN-MIM	0.823 ± 0.045	0.800 ± 0.067	0.695 ± 0.087	0.566 ± 0.109	0.354 ± 0.066	0.7469
PEMix	0.839 ± 0.045	0.844 ± 0.039	0.808 ± 0.043	0.791 ± 0.067	0.822 ± 0.047	**0.8243**

Table 12: TADPOLE: complete macro-F1 results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation.

Model	0	10	20	30	40	50
GNN-Zero	0.825 ± 0.038	0.824 ± 0.033	0.804 ± 0.044	0.804 ± 0.027	0.807 ± 0.019	0.771 ± 0.045
GNN-Median	0.825 ± 0.038	0.836 ± 0.038	0.791 ± 0.030	0.795 ± 0.051	0.801 ± 0.040	0.773 ± 0.038
GNN-MI	0.825 ± 0.038	0.821 ± 0.035	0.811 ± 0.041	0.798 ± 0.032	0.806 ± 0.028	0.763 ± 0.034
FP	0.806 ± 0.065	0.836 ± 0.038	0.808 ± 0.044	0.798 ± 0.037	0.784 ± 0.026	0.782 ± 0.051
PCFI	0.796 ± 0.054	0.810 ± 0.046	0.811 ± 0.038	0.813 ± 0.043	0.804 ± 0.028	0.800 ± 0.035
FISF	0.806 ± 0.040	0.825 ± 0.058	0.811 ± 0.036	0.795 ± 0.045	0.790 ± 0.045	0.797 ± 0.035
GCNmf	0.814 ± 0.043	0.831 ± 0.058	0.792 ± 0.036	0.806 ± 0.042	0.838 ± 0.066	0.812 ± 0.019
GNN-MIM	0.828 ± 0.038	0.838 ± 0.044	0.823 ± 0.049	0.806 ± 0.028	0.809 ± 0.029	0.779 ± 0.058
PEMix	0.828 ± 0.042	0.828 ± 0.049	0.819 ± 0.011	0.843 ± 0.050	0.858 ± 0.034	0.826 ± 0.041

Table 13: TADPOLE: complete macro-F1 results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (higher is better).

Model	60	70	80	90	99	AUC
GNN-Zero	0.758 ± 0.050	0.743 ± 0.067	0.723 ± 0.070	0.723 ± 0.046	0.515 ± 0.250	0.7565
GNN-Median	0.771 ± 0.038	0.761 ± 0.045	0.738 ± 0.036	0.644 ± 0.073	0.241 ± 0.000	0.7398
GNN-MI	0.779 ± 0.049	0.762 ± 0.063	0.744 ± 0.072	0.647 ± 0.107	0.245 ± 0.008	0.7421
FP	0.788 ± 0.044	0.735 ± 0.048	0.770 ± 0.052	0.695 ± 0.060	0.241 ± 0.000	0.7473
PCFI	0.785 ± 0.049	0.774 ± 0.036	0.791 ± 0.050	0.734 ± 0.046	0.241 ± 0.000	0.7593
FISF	0.781 ± 0.060	0.747 ± 0.079	0.773 ± 0.054	0.774 ± 0.061	0.241 ± 0.000	0.7566
GCNmf	0.780 ± 0.016	0.736 ± 0.051	0.798 ± 0.053	0.675 ± 0.073	0.326 ± 0.057	0.7588
GNN-MIM	0.761 ± 0.067	0.752 ± 0.060	0.731 ± 0.058	0.696 ± 0.041	0.241 ± 0.000	0.7482
PEMix	0.816 ± 0.049	0.796 ± 0.041	0.770 ± 0.054	0.817 ± 0.017	0.775 ± 0.057	**0.8094**

Table 14: GraphMSO: complete MAE results under RT (five seeds). Missingness μ is in percent; Rates 0–50%; AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	1.3637 ± 0.0009	1.3732 ± 0.0014	1.3775 ± 0.0016	1.3803 ± 0.0011	1.3835 ± 0.0013	1.3866 ± 0.0011
GNN-MI	1.3604 ± 0.0003	1.3731 ± 0.0014	1.3775 ± 0.0016	1.3801 ± 0.0007	1.3835 ± 0.0013	1.3859 ± 0.0013
GNN-Median	1.3606 ± 0.0008	1.3732 ± 0.0014	1.3774 ± 0.0014	1.3803 ± 0.0012	1.3836 ± 0.0013	1.3859 ± 0.0013
GNN-MIM	1.3602 ± 0.0007	1.3752 ± 0.0020	1.3791 ± 0.0017	1.3818 ± 0.0013	1.3850 ± 0.0018	1.3882 ± 0.0009
FP	1.3583 ± 0.0003	1.3742 ± 0.0012	1.3783 ± 0.0011	1.3814 ± 0.0007	1.3842 ± 0.0006	1.3867 ± 0.0011
PCFI	1.3582 ± 0.0005	1.3729 ± 0.0013	1.3775 ± 0.0011	1.3819 ± 0.0004	1.3853 ± 0.0010	1.3880 ± 0.0013
FISF	1.3632 ± 0.0006	1.3758 ± 0.0012	1.3804 ± 0.0013	1.3849 ± 0.0009	1.3901 ± 0.0013	1.3906 ± 0.0013
GCNmf	1.3583 ± 0.0006	1.3743 ± 0.0013	1.3760 ± 0.0009	1.3775 ± 0.0006	1.3804 ± 0.0012	1.3844 ± 0.0003
PEMix	1.3578 ± 0.0008	1.3715 ± 0.0010	1.3770 ± 0.0010	1.3785 ± 0.0005	1.3813 ± 0.0006	1.3843 ± 0.0013

Table 15: GraphMSO: complete MAE results under RT (five seeds). Missingness μ is in percent; Rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (lower is better).

Model	60	70	80	90	99	AUC
GNN-Zero	1.3872 ± 0.0012	1.3893 ± 0.0008	1.3917 ± 0.0007	1.3941 ± 0.0011	1.3946 ± 0.0003	1.3696
GNN-MI	1.3872 ± 0.0011	1.3894 ± 0.0009	1.3918 ± 0.0006	1.3942 ± 0.0010	1.3949 ± 0.0002	1.3696
GNN-Median	1.3872 ± 0.0011	1.3895 ± 0.0008	1.3918 ± 0.0007	1.3942 ± 0.0009	1.3949 ± 0.0004	1.3697
GNN-MIM	1.3912 ± 0.0023	1.3907 ± 0.0007	1.3935 ± 0.0006	1.3946 ± 0.0005	1.3950 ± 0.0002	1.3712
FP	1.3894 ± 0.0010	1.3925 ± 0.0024	1.3948 ± 0.0009	1.3977 ± 0.0014	1.3964 ± 0.0006	1.3717
PCFI	1.3915 ± 0.0006	1.3960 ± 0.0011	1.3994 ± 0.0009	1.4031 ± 0.0019	1.3976 ± 0.0006	1.3722
FISF	1.3943 ± 0.0008	1.3975 ± 0.0021	1.3976 ± 0.0020	1.3981 ± 0.0010	1.3967 ± 0.0025	1.3738
GCNmf	1.3868 ± 0.0006	1.3889 ± 0.0004	1.3916 ± 0.0005	1.3936 ± 0.0012	1.3945 ± 0.0004	1.3690
PEMix	1.3867 ± 0.0007	1.3887 ± 0.0004	1.3919 ± 0.0006	1.3936 ± 0.0010	1.3944 ± 0.0003	**1.3687**

Table 16: EngRAD: complete MAE results under RT shared-seed missingness (five seeds, mean $\pm$ sample standard deviation). Missingness μ is in percent; rates 0–50%. AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	5.4762 ± 0.0290	5.8960 ± 0.0709	6.3288 ± 0.0976	6.7150 ± 0.0982	7.0647 ± 0.0668	7.4673 ± 0.0701
GNN-MI	5.4364 ± 0.0528	5.7982 ± 0.0400	6.3365 ± 0.0596	6.6810 ± 0.0875	7.0631 ± 0.0963	7.4423 ± 0.0760
GNN-Median	5.4321 ± 0.0377	5.7998 ± 0.0483	6.2912 ± 0.0845	6.7284 ± 0.1058	7.0481 ± 0.0850	7.3866 ± 0.0979
GNN-MIM	5.4174 ± 0.0395	5.8991 ± 0.0407	6.3634 ± 0.1025	6.8770 ± 0.1471	7.4792 ± 0.1344	7.8061 ± 0.0907
FP	5.4181 ± 0.0207	5.4721 ± 0.0229	5.7357 ± 0.1045	5.8706 ± 0.0598	6.2543 ± 0.1555	6.5903 ± 0.2538
PCFI	5.4377 ± 0.0635	5.8268 ± 0.1653	6.3194 ± 0.2407	7.0365 ± 0.1464	7.6782 ± 0.4432	8.4610 ± 0.4943
FISF	5.4015 ± 0.0128	5.5008 ± 0.0281	5.7462 ± 0.0232	5.8516 ± 0.0819	6.0954 ± 0.1340	6.3255 ± 0.1357
GCNmf	5.6397 ± 0.0942	5.9493 ± 0.1136	6.4047 ± 0.1022	6.7618 ± 0.0715	7.1063 ± 0.0705	7.5757 ± 0.1372
PEMix	5.3736 ± 0.0194	5.7791 ± 0.0780	6.1751 ± 0.0635	6.7031 ± 0.0929	7.0101 ± 0.0703	7.1536 ± 0.0951

Table 17: EngRAD: complete MAE results under RT shared-seed missingness (five seeds, mean $\pm$ sample standard deviation). Missingness μ is in percent; rates 60–99%. Trapezoidal AUC is computed over $\mu \in [0, 0.99]$ for each seed and reported as mean $\pm$ sample standard deviation. Lower is better.

Model	60	70	80	90	99	AUC
GNN-Zero	7.8648 ± 0.1059	8.0482 ± 0.1109	8.4417 ± 0.0630	9.0358 ± 0.0620	9.5747 ± 0.0534	7.345719 ± 0.036379
GNN-MI	7.8264 ± 0.0940	8.0547 ± 0.1006	8.4494 ± 0.0871	9.0045 ± 0.0743	9.5532 ± 0.0396	7.322295 ± 0.034537
GNN-Median	7.8062 ± 0.1106	8.0017 ± 0.1033	8.4592 ± 0.0601	9.0617 ± 0.0710	9.6248 ± 0.0392	7.317701 ± 0.046298
GNN-MIM	8.1090 ± 0.0876	8.4326 ± 0.1660	8.9219 ± 0.0680	9.2955 ± 0.0545	9.8190 ± 0.1070	7.584642 ± 0.046850
FP	6.8111 ± 0.1469	6.9421 ± 0.1430	7.2574 ± 0.1736	7.6992 ± 0.1177	9.3221 ± 0.0448	6.515191 ± 0.097710
PCFI	8.9782 ± 0.6702	9.7153 ± 0.6322	10.0908 ± 0.4348	9.4165 ± 0.4403	9.0173 ± 0.0836	7.982851 ± 0.241905
FISF	6.6222 ± 0.1130	6.7754 ± 0.1147	7.1723 ± 0.1351	7.7738 ± 0.1836	10.0132 ± 0.3947	$\mathbf{6.468120 \pm 0.074761}$
GCNmf	7.7186 ± 0.2210	8.0195 ± 0.1684	8.5161 ± 0.0955	8.9464 ± 0.0412	9.7083 ± 0.0460	7.373967 ± 0.043182
PEMix	7.6081 ± 0.2022	7.7819 ± 0.1405	8.3722 ± 0.0844	8.9978 ± 0.1453	9.4479 ± 0.0578	7.206951 ± 0.047572

Table 18: PEMS-BAY: complete MAE results under RT (five seeds). Missingness μ is in percent; Rates 0–50%; AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	3.8743 ± 0.0892	3.8786 ± 0.0594	3.9786 ± 0.0951	4.0962 ± 0.1045	4.3490 ± 0.1824	4.3858 ± 0.3289
GNN-MI	3.9270 ± 0.1181	3.9083 ± 0.0807	3.9255 ± 0.0690	4.0772 ± 0.0740	4.3852 ± 0.1936	4.3559 ± 0.1290
GNN-Median	3.7836 ± 0.0515	3.8937 ± 0.0691	4.0551 ± 0.1404	4.0427 ± 0.0756	4.2222 ± 0.0705	4.6794 ± 0.4123
GNN-MIM	3.8048 ± 0.0902	3.7776 ± 0.0335	3.8063 ± 0.0525	3.9676 ± 0.0963	4.1550 ± 0.1220	4.4299 ± 0.2171
FP	3.7935 ± 0.0726	3.8333 ± 0.0828	3.9304 ± 0.1112	4.0447 ± 0.1111	4.0875 ± 0.1562	4.2809 ± 0.1367
PCFI	3.8095 ± 0.0597	3.9127 ± 0.0745	3.9554 ± 0.0530	4.1170 ± 0.0442	4.3933 ± 0.0864	4.4917 ± 0.0734
FISF	3.8053 ± 0.0601	3.8330 ± 0.0324	4.0053 ± 0.1443	4.0055 ± 0.1086	4.0493 ± 0.0758	4.3540 ± 0.2148
GCNmf	3.7825 ± 0.0618	3.7070 ± 0.0604	3.6741 ± 0.0860	3.7045 ± 0.0971	3.7178 ± 0.1043	3.8272 ± 0.1811
PEMix	3.7540 ± 0.0145	3.7058 ± 0.0441	3.6933 ± 0.0484	3.6723 ± 0.0371	3.6425 ± 0.0474	3.7244 ± 0.0398

Table 19: PEMS-BAY: complete MAE results under RT (five seeds). Missingness μ is in percent; Rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (lower is better).

Model	60	70	80	90	99	AUC
GNN-Zero	4.5305 ± 0.1284	4.7491 ± 0.1542	4.6543 ± 0.1910	4.5401 ± 0.0334	4.6001 ± 0.0208	4.2942
GNN-MI	4.3936 ± 0.0560	4.5632 ± 0.1432	4.5425 ± 0.1582	4.5522 ± 0.0445	4.6196 ± 0.0056	4.2518
GNN-Median	4.5532 ± 0.1190	4.9191 ± 0.2908	4.6654 ± 0.1563	4.5790 ± 0.0628	4.7467 ± 0.2893	4.3408
GNN-MIM	4.4871 ± 0.1917	4.4527 ± 0.1908	4.5787 ± 0.0771	4.5532 ± 0.1272	4.5577 ± 0.0298	4.1934
FP	4.3485 ± 0.1574	4.5187 ± 0.2141	4.5679 ± 0.0786	4.6860 ± 0.1439	4.6140 ± 0.0440	4.2037
PCFI	4.5410 ± 0.1166	4.6474 ± 0.0918	4.5403 ± 0.0502	4.7431 ± 0.0705	4.6168 ± 0.0224	4.3087
FISF	4.4999 ± 0.0960	4.5837 ± 0.1526	4.6479 ± 0.0790	4.6430 ± 0.0494	4.6346 ± 0.0639	4.2378
GCNmf	3.8703 ± 0.1338	4.1159 ± 0.1294	4.3221 ± 0.1160	4.5483 ± 0.1632	4.5958 ± 0.0562	3.9219
PEMix	3.7331 ± 0.0510	3.8074 ± 0.0430	3.9609 ± 0.0684	4.0777 ± 0.1368	4.3424 ± 0.0328	**3.7645**

Table 20: EngRAD: complete MAE results under U-MCAR (mean $\pm$ standard deviation over five seeds). Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	5.462 ± 0.038	5.491 ± 0.047	5.560 ± 0.051	5.610 ± 0.026	5.686 ± 0.050	5.721 ± 0.086
GNN-Median	5.412 ± 0.030	5.496 ± 0.030	5.621 ± 0.046	5.668 ± 0.066	5.773 ± 0.049	5.840 ± 0.036
GNN-MI	5.395 ± 0.037	5.509 ± 0.029	5.589 ± 0.055	5.649 ± 0.049	5.675 ± 0.046	5.721 ± 0.030
FP	5.417 ± 0.028	5.404 ± 0.023	5.434 ± 0.013	5.444 ± 0.021	5.459 ± 0.051	5.429 ± 0.022
PCFI	5.416 ± 0.029	5.604 ± 0.032	5.771 ± 0.026	5.996 ± 0.042	6.062 ± 0.124	6.303 ± 0.090
FISF	5.439 ± 0.053	5.416 ± 0.046	5.453 ± 0.049	5.431 ± 0.035	5.473 ± 0.049	5.499 ± 0.056
GCNmf	5.616 ± 0.040	5.635 ± 0.061	5.895 ± 0.059	5.895 ± 0.057	5.994 ± 0.041	5.816 ± 0.047
GNN-MIM	5.419 ± 0.049	5.536 ± 0.030	5.607 ± 0.051	5.683 ± 0.109	5.723 ± 0.056	5.784 ± 0.093
PEMix	5.370 ± 0.024	5.507 ± 0.046	5.515 ± 0.019	5.564 ± 0.035	5.604 ± 0.028	5.662 ± 0.034

Table 21: EngRAD: complete MAE results under U-MCAR (mean $\pm$ standard deviation over five seeds). Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (lower is better).

Model	60	70	80	90	99	AUC
GNN-Zero	5.842 ± 0.107	5.910 ± 0.036	5.995 ± 0.044	6.277 ± 0.017	7.687 ± 0.076	5.7968
GNN-Median	5.904 ± 0.046	5.989 ± 0.056	6.095 ± 0.040	6.328 ± 0.041	7.769 ± 0.046	5.8600
GNN-MI	5.868 ± 0.115	5.929 ± 0.023	6.111 ± 0.087	6.265 ± 0.025	7.696 ± 0.066	5.8164
FP	5.472 ± 0.028	5.487 ± 0.034	5.545 ± 0.031	5.770 ± 0.087	7.065 ± 0.176	**5.5043**
PCFI	6.408 ± 0.058	6.645 ± 0.079	6.905 ± 0.046	7.407 ± 0.080	9.586 ± 0.077	6.3751
FISF	5.558 ± 0.028	5.637 ± 0.062	5.707 ± 0.046	6.113 ± 0.077	7.422 ± 0.053	5.6040
GCNmf	6.142 ± 0.067	6.107 ± 0.079	6.213 ± 0.124	6.786 ± 0.114	8.117 ± 0.099	6.0604
GNN-MIM	5.801 ± 0.058	5.919 ± 0.200	6.066 ± 0.091	6.334 ± 0.126	7.870 ± 0.086	5.8387
PEMix	5.772 ± 0.055	5.861 ± 0.042	5.980 ± 0.036	6.224 ± 0.029	7.714 ± 0.164	5.7534

Table 22: GraphMSO: complete MAE results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	1.360 ± 0.000	1.373 ± 0.001	1.377 ± 0.001	1.379 ± 0.001	1.381 ± 0.001	1.383 ± 0.001
GNN-Median	1.360 ± 0.001	1.373 ± 0.001	1.377 ± 0.001	1.379 ± 0.001	1.381 ± 0.001	1.383 ± 0.001
GNN-MI	1.363 ± 0.001	1.373 ± 0.001	1.377 ± 0.001	1.379 ± 0.001	1.381 ± 0.001	1.383 ± 0.001
FP	1.358 ± 0.001	1.374 ± 0.001	1.378 ± 0.002	1.380 ± 0.001	1.383 ± 0.001	1.385 ± 0.001
PCFI	1.358 ± 0.001	1.373 ± 0.001	1.377 ± 0.001	1.379 ± 0.001	1.382 ± 0.001	1.384 ± 0.001
FISF	1.363 ± 0.001	1.375 ± 0.001	1.379 ± 0.001	1.381 ± 0.001	1.384 ± 0.001	1.387 ± 0.002
GCNmf	1.358 ± 0.001	1.372 ± 0.001	1.374 ± 0.001	1.376 ± 0.001	1.379 ± 0.001	1.381 ± 0.002
GNN-MIM	1.361 ± 0.001	1.375 ± 0.001	1.378 ± 0.001	1.381 ± 0.001	1.383 ± 0.001	1.384 ± 0.001
PEMix	1.358 ± 0.001	1.371 ± 0.002	1.374 ± 0.001	1.376 ± 0.002	1.378 ± 0.000	1.382 ± 0.001

Table 23: GraphMSO: complete MAE results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (lower is better).

Model	60	70	80	90	99	AUC
GNN-Zero	1.385 ± 0.001	1.387 ± 0.001	1.389 ± 0.001	1.392 ± 0.001	1.396 ± 0.000	1.3684
GNN-Median	1.385 ± 0.001	1.387 ± 0.001	1.389 ± 0.001	1.392 ± 0.001	1.396 ± 0.000	1.3684
GNN-MI	1.385 ± 0.001	1.387 ± 0.001	1.389 ± 0.001	1.392 ± 0.001	1.396 ± 0.001	1.3686
FP	1.387 ± 0.002	1.389 ± 0.001	1.393 ± 0.001	1.395 ± 0.002	1.398 ± 0.001	1.3703
PCFI	1.386 ± 0.002	1.388 ± 0.001	1.393 ± 0.002	1.397 ± 0.002	1.406 ± 0.002	1.3701
FISF	1.388 ± 0.002	1.392 ± 0.001	1.396 ± 0.002	1.401 ± 0.002	1.403 ± 0.002	1.3726
GCNmf	1.384 ± 0.001	1.387 ± 0.001	1.389 ± 0.001	1.392 ± 0.001	1.395 ± 0.000	1.3671
GNN-MIM	1.385 ± 0.002	1.388 ± 0.001	1.390 ± 0.000	1.393 ± 0.001	1.396 ± 0.000	1.3696
PEMix	1.384 ± 0.001	1.386 ± 0.001	1.389 ± 0.001	1.392 ± 0.001	1.396 ± 0.001	**1.3669**

Table 24: PEMS-BAY: complete MAE results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation. Lower is better.

Model	0	10	20	30	40	50
GNN-Zero	3.852 ± 0.074	3.901 ± 0.070	3.882 ± 0.058	4.060 ± 0.132	4.184 ± 0.194	4.201 ± 0.088
GNN-Median	3.832 ± 0.049	3.833 ± 0.045	3.907 ± 0.034	3.937 ± 0.042	4.030 ± 0.123	4.081 ± 0.037
GNN-MI	3.895 ± 0.095	3.801 ± 0.031	3.914 ± 0.141	4.068 ± 0.084	4.088 ± 0.062	4.262 ± 0.074
FP	3.799 ± 0.079	3.862 ± 0.024	3.968 ± 0.040	3.925 ± 0.049	3.975 ± 0.054	3.982 ± 0.035
PCFI	3.797 ± 0.075	3.905 ± 0.093	3.981 ± 0.096	4.021 ± 0.141	4.053 ± 0.065	4.089 ± 0.041
FISF	3.814 ± 0.053	3.804 ± 0.020	3.922 ± 0.024	3.857 ± 0.021	3.969 ± 0.037	4.041 ± 0.063
GCNmf	3.776 ± 0.068	3.316 ± 0.058	3.278 ± 0.027	3.266 ± 0.026	3.294 ± 0.034	3.279 ± 0.025
GNN-MIM	3.830 ± 0.065	3.815 ± 0.073	3.894 ± 0.103	3.928 ± 0.029	3.849 ± 0.041	3.933 ± 0.046
PEMix	3.753 ± 0.010	3.424 ± 0.027	3.395 ± 0.021	3.357 ± 0.021	3.390 ± 0.024	3.435 ± 0.051

Table 25: PEMS-BAY: complete MAE results under U-MCAR (mean ± standard deviation over five seeds). Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (lower is better).

Model	60	70	80	90	99	AUC
GNN-Zero	4.155 ± 0.058	4.104 ± 0.135	4.124 ± 0.131	4.215 ± 0.166	4.574 ± 0.049	4.0600
GNN-Median	4.248 ± 0.115	4.037 ± 0.091	4.220 ± 0.087	4.386 ± 0.086	4.474 ± 0.068	4.0389
GNN-MI	4.101 ± 0.084	4.300 ± 0.056	4.359 ± 0.153	4.138 ± 0.120	4.500 ± 0.059	4.0795
FP	4.108 ± 0.102	4.148 ± 0.055	4.290 ± 0.078	4.380 ± 0.057	4.527 ± 0.012	4.0356
PCFI	4.154 ± 0.061	4.260 ± 0.097	4.395 ± 0.067	4.518 ± 0.108	4.677 ± 0.047	4.1154
FISF	4.075 ± 0.035	4.228 ± 0.064	4.414 ± 0.112	4.608 ± 0.167	4.708 ± 0.055	4.0713
GCNmf	3.333 ± 0.038	3.375 ± 0.011	3.497 ± 0.021	3.656 ± 0.038	4.299 ± 0.054	**3.3935**
GNN-MIM	3.956 ± 0.030	3.977 ± 0.048	4.004 ± 0.066	4.047 ± 0.020	4.386 ± 0.058	3.9091
PEMix	3.437 ± 0.029	3.477 ± 0.052	3.564 ± 0.076	3.672 ± 0.094	3.988 ± 0.046	3.4638

E Implementation and Reproducibility Details

This appendix records optimization, covariance assumptions, hyperparameter ranges, and computational complexity. These details are separated from the conceptual method in Section 4 so the main text follows the mechanism-to-model argument.

Fully-missing nodes. When $\mathcal{O}_i = \emptyset$, the feature contribution to the responsibility vanishes, while the positional term remains active. Consequently, the imputation can remain node-specific when the PE-conditioned responsibilities distinguish nodes.

Learning the mixture. The mixture parameters $(\pi_k, \mu_k, \sigma_k^2)$ are *not* fixed by a separate EM fit; they are trainable and optimised end-to-end with the GCN weights $W^{(1)}, W^{(2)}$ by minimising the cross-entropy of the node-classification loss. We initialise them with a few EM iterations on the mean-imputed extended matrix $Z = [\widetilde{X} \mid \mathrm{PE}]$, which gives a sensible starting partition, and then let backpropagation refine them jointly with the classifier. The mixing weights are kept on the simplex by an unconstrained softmax re-parameterisation, and the variances are kept positive by optimising $\log \sigma_k^2$.

Diagonal covariances. We use diagonal covariances $\mathrm{diag}(\sigma_k^2)$ throughout. This is what makes the marginalisation of the missing coordinates closed-form and the expected ReLU in the probabilistic first layer analytic; full covariances would couple observed and missing coordinates and require more expensive conditional-moment computations, although Gaussian conditioning and scalar expected ReLUs remain analytic. The diagonal assumption is the main modelling limitation of PEMix.

Hyperparameters. For all competing methods, we follow the hyperparameter search spaces and model-selection procedures of Ferrini et al. (2026) and of the original methods, using the official implementations when available. Model selection is performed exclusively on the validation set, without access to test performance.

For the controlled-missingness experiments, hyperparameter selection is performed independently for RT and U-MCAR. Within each mechanism, candidate configurations are evaluated over the full missingness curve and selected according to validation AUC. The selected network and optimization hyperparameters are reported for TADPOLE, EngRAD, GraphMSO, and PEMS-BAY in Tables 26, 27, 28, and 29, respectively.

For datasets with naturally occurring missingness, no additional missingness is injected and hyperparameters are selected directly according to validation MAE. The selected configurations for AQI, PV-US, and METR-LA are reported in Tables 30, 31, and 32, respectively.

For PEMix and GCNmf, the GMM-specific hyperparameters are reported separately. In the TADPOLE experiments, PEMix uses $K = 5$ mixture components and PE dimension $q = 8$, fixed across both missingness mechanisms and all missingness rates. For EngRAD, GraphMSO, and PEMS-BAY, the selected GMM configurations are reported in Tables 33, 34, and 35, respectively. The corresponding settings for AQI, PV-US, and METR-LA are reported in Tables 36, 37, and 38.

Table 26: Hyperparameter configurations selected for TADPOLE. Hyperparameters are selected separately for RT and U-MCAR using validation performance, and the selected configuration is then kept fixed across all missingness rates within each mechanism.

Mechanism	Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
RT	GNN-Zero	128	0.5	0.01	0	500	50
RT	GNN-Median	128	0.5	0.01	0	500	50
RT	GNN-MI	128	0.5	0.01	0	500	50
RT	GNN-MIM	128	0.5	0.01	0	500	50
RT	FP	64	0.5	0.01	0	1000	40
RT	PCFI	64	0.5	0.01	0	1000	40
RT	FISF	64	0.5	0.01	0	1000	40
RT	GCNmf	16	0	0.01	10^{-5}	1000	40
RT	PEMix	16	0	0.005	0	500	50
U-MCAR	GNN-Zero	128	0.5	0.01	0	500	50
U-MCAR	GNN-Median	128	0.5	0.01	0	500	50
U-MCAR	GNN-MI	128	0.5	0.01	0	500	50
U-MCAR	GNN-MIM	128	0.5	0.01	0	500	50
U-MCAR	FP	64	0.5	0.01	0	1000	40
U-MCAR	PCFI	64	0.5	0.01	0	1000	40
U-MCAR	FISF	64	0.5	0.01	0	1000	40
U-MCAR	GCNmf	16	0	0.01	10^{-5}	1000	40
U-MCAR	PEMix	16	0	0.005	0	500	50

Table 27: Hyperparameter configurations selected for EngRAD. Hyperparameters are selected separately for RT and U-MCAR using validation performance, and the selected configuration is then kept fixed across all missingness rates within each mechanism.

Mechanism	Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
RT	GNN-Zero	128	0.5	0.001	0	500	50
RT	GNN-Median	128	0.3	0.001	0	500	50
RT	GNN-MI	128	0.5	0.001	0	500	50
RT	GNN-MIM	128	0.3	0.001	0	500	50
RT	FP	64	0.3	0.001	0	1000	40
RT	PCFI	64	0.5	0.001	0	1000	40
RT	FISF	64	0.5	0.01	0	1000	40
RT	GCNmf	16	0.5	0.003	0	1000	40
RT	PEMix	32	0.3	0.001	0	500	50
U-MCAR	GNN-Zero	128	0.5	0.001	0	500	50
U-MCAR	GNN-Median	128	0.3	0.001	0	500	50
U-MCAR	GNN-MI	128	0.5	0.001	0	500	50
U-MCAR	GNN-MIM	128	0.1	0.001	0	500	50
U-MCAR	FP	64	0.5	0.003	0	1000	40
U-MCAR	PCFI	64	0.5	0.001	0	1000	40
U-MCAR	FISF	64	0.5	0.001	0	1000	40
U-MCAR	GCNmf	16	0.5	0.003	0	1000	40
U-MCAR	PEMix	16	0.5	0.001	0	500	50

Table 28: Hyperparameter configurations selected for GraphMSO. Hyperparameters are selected separately for RT and U-MCAR using validation performance, and the selected configuration is then kept fixed across all missingness rates within each mechanism.

Mechanism	Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
RT	GNN-Zero	128	0.5	0.003	0	500	50
RT	GNN-Median	128	0.5	0.01	0	500	50
RT	GNN-MI	128	0.5	0.003	0	500	50
RT	GNN-MIM	128	0.1	0.003	0	500	50
RT	FP	64	0.5	0.01	0	1000	40
RT	PCFI	64	0.5	0.01	0	1000	40
RT	FISF	64	0.5	0.01	0	1000	40
RT	GCNmf	16	0.5	0.01	0	1000	40
RT	PEMix	16	0.5	0.01	0	500	50
U-MCAR	GNN-Zero	128	0.5	0.003	0	500	50
U-MCAR	GNN-Median	128	0.5	0.01	0	500	50
U-MCAR	GNN-MI	128	0.5	0.01	0	500	50
U-MCAR	GNN-MIM	128	0.3	0.003	0	500	50
U-MCAR	FP	64	0.5	0.01	0	1000	40
U-MCAR	PCFI	64	0.5	0.01	0	1000	40
U-MCAR	FISF	64	0.5	0.01	0	1000	40
U-MCAR	GCNmf	16	0	0.01	0	1000	40
U-MCAR	PEMix	16	0.5	0.01	0	500	50

Table 29: Hyperparameter configurations selected for PEMS-BAY. Hyperparameters are selected separately for RT and U-MCAR using validation performance, and the selected configuration is then kept fixed across all missingness rates within each mechanism.

Mechanism	Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
RT	GNN-Zero	128	0	0.001	0	500	50
RT	GNN-Median	128	0	0.003	0	500	50
RT	GNN-MI	128	0.3	0.001	0	500	50
RT	GNN-MIM	128	0	0.001	0	500	50
RT	FP	64	0.3	0.003	0	1000	40
RT	PCFI	64	0.1	0.003	0	1000	40
RT	FISF	64	0.5	0.001	0	1000	40
RT	GCNmf	16	0	0.01	0	1000	40
RT	PEMix	16	0.5	0.01	0	500	50
U-MCAR	GNN-Zero	128	0.5	0.003	0	500	50
U-MCAR	GNN-Median	128	0.1	0.001	0	500	50
U-MCAR	GNN-MI	128	0.5	0.003	0	500	50
U-MCAR	GNN-MIM	128	0	0.001	0	500	50
U-MCAR	FP	64	0	0.003	0	1000	40
U-MCAR	PCFI	64	0.1	0.001	0	1000	40
U-MCAR	FISF	64	0.1	0.001	0	1000	40
U-MCAR	GCNmf	16	0	0.01	0	1000	40
U-MCAR	PEMix	16	0.5	0.001	0	500	50

Table 30: Hyperparameter configurations selected for AQI under natural missingness. Configurations are selected using validation MAE; no additional missingness is injected. GMM settings are reported separately.

Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
GNN-Zero	128	0.1	0.01	0	500	50
GNN-Median	128	0.3	0.01	0	500	50
GNN-MI	128	0.5	0.01	0	500	50
GNN-MIM	128	0.5	0.01	0	500	50
FP	64	0.3	0.001	0	1000	40
PCFI	64	0.5	0.003	0	1000	40
FISF	64	0.1	0.01	0	1000	40
GCNmf	16	0	0.01	0	1000	40
PEMix	16	0.5	0.003	0	500	50

Table 31: Hyperparameter configurations selected for PV-US under natural missingness. Configurations are selected using validation MAE; no additional missingness is injected. GMM settings are reported separately.

Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
GNN-Zero	128	0.5	0.003	0	500	50
GNN-Median	128	0	0.003	0	500	50
GNN-MI	128	0.3	0.003	0	500	50
GNN-MIM	128	0.5	0.001	0	500	50
FP	64	0.3	0.003	0	1000	40
PCFI	64	0.5	0.003	0	1000	40
FISF	64	0.3	0.001	0	1000	40
GCNmf	16	0	0.01	0	1000	40
PEMix	16	0.5	0.01	0	500	50

Table 32: Hyperparameter configurations selected for METR-LA under natural missingness. Configurations are selected using validation MAE; no additional missingness is injected. GMM settings are reported separately.

Method	Hidden	Dropout	LR	Weight decay	Max epochs	Patience
GNN-Zero	128	0	0.001	0	500	50
GNN-Median	128	0.1	0.001	0	500	50
GNN-MI	128	0	0.001	0	500	50
GNN-MIM	128	0.3	0.003	0	500	50
FP	64	0	0.003	0	1000	40
PCFI	64	0	0.003	0	1000	40
FISF	64	0	0.003	0	1000	40
GCNmf	16	0	0.001	0	1000	40
PEMix	16	0.5	0.01	0	500	50

Table 33: GMM settings selected for EngRAD. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Mechanism	Method	K	q	Likelihood
RT	GCNmf	5	0	legacy
RT	PEMix	8	16	gaussian
U-MCAR	GCNmf	5	0	legacy
U-MCAR	PEMix	5	8	gaussian

Table 34: GMM settings selected for GraphMSO. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Mechanism	Method	K	q	Likelihood
RT	GCNmf	5	0	legacy
RT	PEMix	3	16	gaussian
U-MCAR	GCNmf	5	0	gaussian
U-MCAR	PEMix	2	8	gaussian

Table 35: GMM settings selected for PEMS-BAY. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Mechanism	Method	K	q	Likelihood
RT	GCNmf	5	0	legacy
RT	PEMix	5	16	gaussian
U-MCAR	GCNmf	5	0	legacy
U-MCAR	PEMix	5	8	gaussian

Table 36: GMM settings selected for AQI. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Method	K	q	Likelihood
GCNmf	5	0	legacy
PEMix	5	8	gaussian

Table 37: GMM settings selected for PV-US. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Method	K	q	Likelihood
GCNmf	5	0	gaussian
PEMix	3	4	gaussian

Table 38: GMM settings selected for METR-LA. K is the number of mixture components and q is the PE dimension; GCNmf does not use PE ($q = 0$).

Method	K	q	Likelihood
GCNmf	5	0	gaussian
PEMix	5	16	gaussian

E.1 Computational Complexity

Notation and conventions. Let n be the number of nodes, $m = \mathrm{nnz}(\hat{\mathbf{A}}) \geq n$ the number of nonzeros of the propagation operator (both edge directions and self-loops), d the input dimension, h the hidden dimension, c the output dimension, K the number of mixture components, q the PE dimension, and R the number of training epochs. All costs refer to a single input graph with missingness mask $\mathbf{M}$. All models share the same backbone. For a two-layer GCN that applies the linear map before propagation, one full-batch forward (and, up to a constant, backward) pass costs

$$\mathcal{C}_{\mathrm{B}}(d) = O(ndh + mh + nhc + mc). \tag{14}$$

We report separately (i) the cost of handling missing features, which for all baselines depends on $\mathbf{M}$ and must be recomputed whenever the mask changes, and (ii) the per-epoch cost of the learned model.

Imputation baselines. Zero, mean, and median imputation require one pass over the feature matrix, $O(nd)$ (median via linear-time selection; sorting-based implementations cost $O(nd \log n)$). GNN-MIM additionally concatenates the mask, so its first layer acts on $2d$ inputs and costs $O(2ndh + mh)$; only the feature-transform term doubles. FP performs T_{FP} sparse propagations, each followed by re-injection of the observed values, for $O(T_{\mathrm{FP}}\, dm)$ time and $O(m + nd)$ memory. PCFI additionally computes channel-wise shortest-path distances to observed nodes ($O(dm)$ via multi-source BFS), diffuses each channel with a channel-specific confidence-weighted operator ($O(T_{\mathrm{PC}}\, dm)$ time and $O(dm)$ memory when the d operators are materialized), and applies one inter-channel step based on the $d \times d$ channel correlation matrix, which costs $O(nd^2)$. FISF runs a pre-diffusion on all channels (T_1 iterations), selects the ρd channels with the lowest variance, and re-diffuses them (T_2 iterations) after recomputing their distances, giving $O((T_1 + \rho T_2)\, dm)$. After imputation, training costs $R\,\mathcal{C}_{\mathrm{B}}(d)$.

Mixture-based models. GCNmf and PEMix handle missing features inside every forward pass. For each component k, the first layer propagates the component-wise mean and variance matrices through $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}} \odot \hat{\mathbf{A}}$, evaluates the expected ReLU, and mixes the results with the responsibilities γ_{ik}. The per-epoch cost is

$$\mathcal{C}_{\mathrm{mix}} = O\big(K(ndh{+}mh){+}nK(d{+}q)\big){+}O(nhc{+}mc) = O\big(K(ndh{+}mh){+}nKq\big){+}O(nhc{+}mc), \tag{15}$$

with $q = 0$ for GCNmf. The positional encodings enter only the responsibilities and are never propagated, so $\mathbf{W}^{(1)} \in \mathbb{R}^{d \times h}$ and the message-passing cost is identical to GCNmf. The additional term $O(nKq)$ is dominated by $O(Kndh)$ whenever $q \leq dh$, which holds in all our experiments; PEMix therefore has the same per-epoch complexity as GCNmf. Both are approximately K times more expensive than the backbone's first layer and store $O(Knh)$ first-layer activations for backpropagation. Mixture parameters are initialized by I EM iterations on $[\tilde{\mathbf{X}} \| \mathbf{P}]$, costing $O(I\, nK(d + q))$ for diagonal covariances.

Positional encodings. PEMix computes the $q{+}1$ smallest eigenpairs of the normalized Laplacian with a sparse iterative eigensolver. Each Lanczos/LOBPCG iteration costs $O(mq + nq^2)$, while the number of iterations depends on the tolerance and on the spectral gap, so we do not state a closed-form bound. This computation depends only on $\mathbf{A}$: it is performed once per graph and shared across all missingness rates, seeds, and masks. In contrast, the preprocessing of the baselines is recomputed for every mask.

PEMix additionally requires one sparse partial eigendecomposition per graph, independent of the mask.

Inference. For a fixed mask, imputation-based baselines reuse the completed matrix and require one backbone forward pass, whereas GCNmf and PEMix require one mixture forward pass. When the mask changes at test time, the baselines must re-run their imputation, while the mixture models handle the new mask within the forward pass.

Summary. Structural conditioning comes at a small cost: PEMix has the same per-epoch complexity as GCNmf, adding only a dominated $O(nKq)$ term to the mixture responsibilities and a

Table 39: Computational cost per input graph. Mask-dependent costs are recomputed for every missingness pattern. †: $O(nd \log n)$ with sorting. ‡: implementation-dependent (see text).

Method	Mask-dependent one-off	Per epoch	Extra memory
GNN-Zero / Mean	$O(nd)$	$\mathcal{C}_{\mathrm{B}}(d)$	–
GNN-Median	$O(nd)^{\dagger}$	$\mathcal{C}_{\mathrm{B}}(d)$	–
GNN-MIM	$O(nd)$	$\mathcal{C}_{\mathrm{B}}(d) + O(ndh)$	$O(nd)$
FP	$O(T_{\mathrm{FP}}\, dm)$	$\mathcal{C}_{\mathrm{B}}(d)$	$O(nd)$
PCFI	$O(T_{\mathrm{PC}}\, dm + nd^2)^{\ddagger}$	$\mathcal{C}_{\mathrm{B}}(d)$	$O(dm + d^2)^{\ddagger}$
FISF	$O((T_1 + \rho T_2)\, dm)^{\ddagger}$	$\mathcal{C}_{\mathrm{B}}(d)$	$O(dm)^{\ddagger}$
GCNmf	$O(I\, nKd)$	$O(K(ndh + mh)) + O(nhc + mc)$	$O(Knh)$
PEMix	$O(I\, nK(d + q))$	$O(K(ndh + mh) + nKq) + O(nhc + mc)$	$O(Knh + nq)$

single sparse partial eigendecomposition per graph. Compared with imputation-based baselines, both mixture models incur a moderate constant-factor overhead per epoch, proportional to the small number of mixture components K used in our experiments", because missing features are handled inside the forward pass. In exchange, they require no mask-dependent preprocessing, avoid the $O(nd^2)$ inter-channel cost and $O(dm)$ channel-specific operators of PCFI and FISF, and handle new missingness patterns without re-running imputation.

F Ablations

This appendix complements RQ3 (Table 4). We report a matched ablation that removes PE conditioning from PEMix, PE-only controls, the effect of jointly optimising the mixture, and the complete per-level results of the positional-encoding ablation on TADPOLE. We additionally include an MLP PE-only control to separate PE information from message passing

F.1 Matched structural-conditioning ablation

Most models in Table 4 differ from PEMix not only in how they use PE but also in their architecture and optimisation settings. To isolate the contribution of PE conditioning alone, we compare PEMix with the same model in which the positional encodings are set to zero, so that the mixture responsibilities receive no structural information. This variant is a GCNmf trained exactly as PEMix. The two models share masks, splits, optimisation settings, and the probabilistic first layer, and coincide exactly at $\mu = 0$, where no feature is missing. As shown in Figure 3, removing PE conditioning lowers the AUC from 0.8243 to 0.7330 on TADPOLE where the gap is largest at high missingness: from $\mu = 0.6$ onwards it grows steadily, reaching 0.49.

F.2 PE-only controls

On TADPOLE, the PE-only row in Table 4 is a two-layer GCN whose only node input is the PE, with the same depth, hidden width, dropout, and learning rate as PEMix. To separate the contribution of PE from that of message passing, we also train an MLP with the same depth, width, dropout, and learning rate on the PE alone, which ignores the graph. Since neither control receives the node features, its score does not depend on μ. The GCN control obtains 0.295 ± 0.092 macro-F1 and the MLP control 0.254 ± 0.027, both well below PEMix at every missingness level, including $\mu = 0.99$ (0.822 ± 0.047).

F.3 Effect of joint GMM optimisation

Following Taguchi et al. (2021), we initialise the mixture parameters with EM and subsequently optimise them jointly with the downstream prediction objective. To assess the effect of this design choice, we compare the full model against a *frozen-GMM* variant, in which the mixture means, variances, and weights are fixed after EM initialisation, while the GNN and prediction head remain trainable.

We evaluate both variants on EngRAD under RT at missing rates of 20%, 50%, and 80%, using five seeds. As shown in Table 40, jointly optimising the mixture consistently improves downstream forecasting performance compared with keeping the EM-initialised mixture fixed.

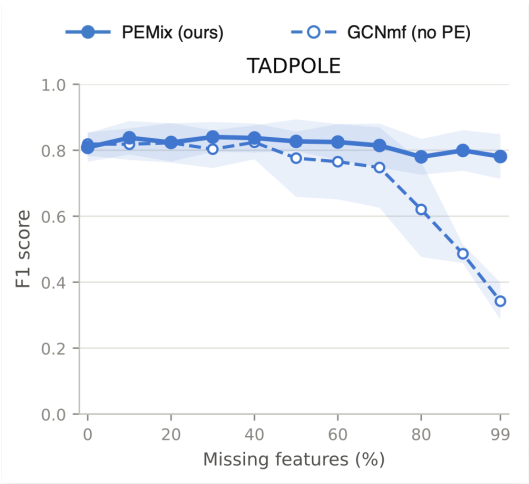


Figure 3: Matched ablation between PEMix and the same probabilistic recovery model without PE conditioning, obtained by setting the positional encodings to zero. The models share masks, splits, optimization settings, and the probabilistic first layer; at $\mu = 0$ they coincide exactly. Mean $\pm$ standard deviation over five seeds.

Table 40: Effect of jointly optimising the GMM on EngRAD under RT. Values are mean $\pm$ standard deviation over five seeds. Lower is better.

Missing rate	Joint GMM	Frozen GMM
20%	$\mathbf{6.1751 \pm 0.0635}$	6.287 ± 0.086
50%	$\mathbf{7.1536 \pm 0.0951}$	7.364 ± 0.113
80%	$\mathbf{8.3722 \pm 0.0844}$	8.523 ± 0.080

F.4 Positional-encoding ablation: complete results on TADPOLE

Tables 41 and 42 report, for every model in Table 4, TADPOLE macro-F1 (mean $\pm$ standard deviation over five seeds) at each missingness rate under RT. The last column gives the AUC reported in Table 4. GCNmf + PE (downstream) is a latent-fusion variant: the mixture is fitted on the node features alone, and PE are concatenated to the representation produced by the probabilistic first layer, so they cannot affect the mixture responsibilities.

Table 41: TADPOLE: positional-encoding ablation under RT, macro-F1 (mean ± standard deviation over five seeds). "Mechanism" uses PE while handling missing features; "Downstream" concatenates PE after missing-feature handling. Missingness μ is in percent; rates 0–50%; AUC is reported in the continuation.

PE usage	Model	0	10	20	30	40	50
Mechanism	PEMix	0.828 ± 0.042	0.864 ± 0.041	0.846 ± 0.050	0.838 ± 0.050	0.844 ± 0.054	0.826 ± 0.028
	FP + PE	0.806 ± 0.065	0.818 ± 0.024	0.805 ± 0.036	0.803 ± 0.030	0.812 ± 0.055	0.776 ± 0.074
	PCFI + PE	0.796 ± 0.054	0.790 ± 0.047	0.814 ± 0.036	0.817 ± 0.040	0.812 ± 0.041	0.818 ± 0.026
	FISF + PE	0.806 ± 0.040	0.812 ± 0.026	0.806 ± 0.042	0.815 ± 0.047	0.813 ± 0.041	0.803 ± 0.048
Downstream	GCNmf + PE	0.832 ± 0.055	0.819 ± 0.053	0.810 ± 0.050	0.812 ± 0.046	0.854 ± 0.035	0.816 ± 0.037
	FP + PE	0.842 ± 0.017	0.824 ± 0.026	0.831 ± 0.030	0.785 ± 0.047	0.805 ± 0.051	0.780 ± 0.054
	PCFI + PE	0.841 ± 0.024	0.809 ± 0.032	0.815 ± 0.034	0.811 ± 0.017	0.807 ± 0.036	0.824 ± 0.038
	FISF + PE	0.835 ± 0.016	0.816 ± 0.016	0.810 ± 0.055	0.810 ± 0.042	0.845 ± 0.033	0.765 ± 0.063
–	PE only	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092

Table 42: TADPOLE: positional-encoding ablation under RT, macro-F1 (mean ± standard deviation over five seeds). "Mechanism" uses PE while handling missing features; "Downstream" concatenates PE after missing-feature handling. Missingness μ is in percent; rates 60–99% and trapezoidal AUC over $\mu \in [0, 0.99]$ (higher is better).

PE usage	Model	60	70	80	90	99	AUC
Mechanism	PEMix	0.839 ± 0.045	0.844 ± 0.039	0.808 ± 0.043	0.791 ± 0.067	0.822 ± 0.047	**0.8243**
	FP + PE	0.792 ± 0.067	0.722 ± 0.146	0.685 ± 0.125	0.562 ± 0.082	0.412 ± 0.075	0.7336
	PCFI + PE	0.818 ± 0.029	0.796 ± 0.061	0.670 ± 0.119	0.559 ± 0.114	0.464 ± 0.076	0.7474
	FISF + PE	0.822 ± 0.046	0.777 ± 0.102	0.701 ± 0.041	0.633 ± 0.105	0.491 ± 0.057	0.7573
Downstream	GCNmf + PE	0.755 ± 0.079	0.734 ± 0.163	0.577 ± 0.131	0.490 ± 0.035	0.339 ± 0.073	0.7211
	FP + PE	0.782 ± 0.078	0.775 ± 0.048	0.744 ± 0.133	0.762 ± 0.054	0.631 ± 0.114	0.7754
	PCFI + PE	0.763 ± 0.099	0.807 ± 0.063	0.719 ± 0.126	0.736 ± 0.095	0.600 ± 0.069	0.7745
	FISF + PE	0.825 ± 0.052	0.807 ± 0.053	0.731 ± 0.153	0.737 ± 0.157	0.603 ± 0.092	0.7798
–	PE only	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.295 ± 0.092	0.2923