RETHINKING EPISTEMIC UNCERTAINTY IN NODE CLASSIFICATION THROUGH INFORMATION GROWTH

Anonymous authors
Paper under double-blind review

ABSTRACT

Epistemic uncertainty should decrease as additional information about the data-generating process (DGP) becomes available to the predictor. Yet, existing graph evidential deep learning (EDL) methods for node classification typically construct epistemic uncertainty from graph-specific properties and evaluate it on downstream tasks such as out-of-distribution detection, which do not test its reducibility as information about the DGP increases. To make reducibility directly testable, we introduce a statistical framework for studying epistemic uncertainty under information growth. Our framework specifies an information-growth experimental protocol and a consistency criterion for epistemic predictors, while using projective graph DGPs to ensure that growing graphs, which in general need not provide increasing information about the same DGP, constitute coherent observations of the same underlying process. We show that EDL methods do not explicitly estimate data uncertainty arising from a single finite graph observation and instead regulate epistemic uncertainty through model hyperparameters, precluding consistency, as corroborated by controlled information-growth experiments. As an alternative, we propose graph bootstrap ensembles, capturing both data and procedural uncertainty through graph resampling and randomized training. Under the same experimental protocol, these ensembles exhibit epistemic uncertainty reduction beyond standard deep ensembles. These findings support bootstrap ensembles as candidate consistent epistemic predictors under information growth.

1 INTRODUCTION

Machine learning models are increasingly expected to provide not only accurate predictions, but also reliable estimates of predictive uncertainty (Abdar et al., 2021; Gawlikowski et al., 2023; Wang et al., 2025). Predictive uncertainty is commonly decomposed into *aleatoric* uncertainty, arising from intrinsic ambiguity in the data-generating process (DGP) and thus irreducible, and *epistemic* uncertainty, arising from limited knowledge of that process and thus reducible with additional data (Hüllermeier & Waegeman, 2021; Sale et al., 2026). In classical supervised learning with independent and identically distributed (i.i.d.) observations, under suitable regularity conditions, epistemic uncertainty is operationally characterized to decrease monotonically as information increases and eventually vanish (Bengs et al., 2022; Juergens et al., 2024; Shen et al., 2024).

Extending this characterization to node classification on graphs is non-trivial. While increasing sample size provides a natural information-growth mechanism in the i.i.d. setting, no analogous mechanism is readily available for node classification. Intuitively, information growth would involve progressively larger graph observations and training sets becoming available, allowing predictors to be retrained on more informative observations for the same node classification problem. However, nodes are dependent, and observing more of them does not necessarily yield more information about the same underlying DGP (Shalizi & Rinaldo, 2013). Consequently, node classification currently lacks a framework for assessing the reducibility of epistemic uncertainty under information growth.

Graph uncertainty quantification has been studied using a variety of approaches (see Wang et al. (2024); Chen et al. (2024) for reviews). Among these, evidential deep learning (EDL) methods (Zhao et al., 2020; Stadler et al., 2021; Hart et al., 2023; Damke & Hüllermeier, 2024a;b) are particularly relevant because they make the epistemic interpretation a central modeling objective, prescribing appropriate uncertainty behavior through graph-specific principles, including reducibil-

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.

ity through *feature support* and *structural support*, and the reducibility or irreducibility due to *neighborhood conflict*. However, these principles are defined with respect to the features and topology of a fixed observed graph rather than under information growth. Empirical evaluations likewise assess epistemic uncertainty through its utility for out-of-distribution (OOD) detection, whose optimal predictors need not coincide with predictors of reducible epistemic uncertainty (Paplhám et al., 2026). While informative about task-specific downstream utility, such evaluations therefore cannot establish whether an uncertainty score has this epistemic interpretation.

Together, these limitations raise fundamental questions about how epistemic uncertainty should be defined and modeled in node classification. We investigate these issues through three research questions: **(RQ1)** What constitutes a principled statistical framework for epistemic predictors in node classification? **(RQ2)** Are graph EDL models consistent epistemic predictors, and, if not, what are their fundamental limitations? **(RQ3)** Which alternative approach can address these limitations?

Contributions. We address these questions through three main contributions.

(1) We establish the first statistical framework for epistemic uncertainty under information growth in node classification, comprising an information-growth experimental protocol and a definition of consistent epistemic predictors. The former naturally requires *projective* graph DGPs to preserve the underlying node-classification problem across nested graphs and learning tasks (Shalizi & Rinaldo, 2013), while the latter requires decreasing and asymptotically vanishing uncertainty. Together, these provide a statistical basis for assessing epistemic reducibility beyond conventional evaluations.

(2) We unify graph EDL objectives and characterize their population optima, showing how model hyperparameters artificially regulate epistemic uncertainty in the absence of explicit estimation of data uncertainty (Shen et al., 2024), arising from observing a single finite realization of the DGP, ultimately precluding epistemic consistency. Controlled information-growth experiments on projective graphs confirm insufficient epistemic concentration and persistent parameterization dependence, corroborating the structural limitations that undermine their use as consistent epistemic predictors.

(3) Motivated by evidence that bootstrap-based ensembles capture data and procedural uncertainty (Jimenez et al., 2026), thus enabling decreasing epistemic uncertainty in the i.i.d. setting (Jain & Bates, 2025), we extend this approach to graphs through *graph bootstrap ensembles*. Under the same protocol as in (2), the experiments support these ensembles as a family of candidate consistent epistemic predictors, demonstrating the ability to recover epistemic concentration absent from standard deep ensembles.

Together, these contributions establish a statistical foundation for understanding epistemic uncertainty in node classification, reveal the limitations of existing evidential approaches when assessed through information growth, and introduce graph bootstrap ensembles as a theoretically motivated, empirically supported alternative for modeling epistemic uncertainty that contracts under information growth.

2 Background on Predictive Uncertainty and Motivation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected attributed graph with $|\mathcal{V}| = N$, adjacency matrix $A \in \{0,1\}^{N\times N}$, and node feature matrix $X \in \mathbb{R}^{N\times D}$. Let $\mathcal{Y} = \{1, \dots, C\}$ denote the set of class labels. In semi-supervised node classification, a set of labeled nodes $\mathcal{T}^{(\mathrm{tr})} \subset \mathcal{V}$ is used to predict the labels of the remaining nodes $\mathcal{T}^{(\mathrm{te})} = \mathcal{V} \setminus \mathcal{T}^{(\mathrm{tr})}$. We consider quantifying the uncertainty of these predictions. A probabilistic classifier associates each node $i \in \mathcal{V}$ with a predictive categorical distribution $\theta^{(i)} \in \Delta_C$, where $\Delta_C = \{\theta \in \mathbb{R}^C : \theta_c \geq 0,\ \sum_{c=1}^{C} \theta_c = 1\}$ denotes the probability simplex over $\mathcal{Y}$. We adopt a second-order probabilistic framework in which the predictive distribution is itself regarded as a random variable $\Theta^{(i)}$. The second-order predictor $Q^{(i)}$ specifies a distribution over predictive distributions $\Theta^{(i)} \sim Q^{(i)}$, with $Q : (\mathcal{G}, i) \mapsto \mathcal{P}(\Delta_C)$. We use the standard information-theoretic decomposition of total predictive uncertainty into aleatoric and epistemic components (Depeweg, 2019; Wimmer et al., 2023). Since the following quantities are defined node-wise, we omit the superscript $^{(i)}$ and define the predictive mean by $\bar{\theta} = \mathbb{E}_Q[\Theta]$. The total uncertainty (TU) is the

Shannon entropy of the predictive mean,

$$\mathrm{TU}(Q) = H(\mathbb{E}_Q[\Theta]) = -\sum_{c=1}^{C} \bar{\theta}_c \log \bar{\theta}_c. \tag{1}$$

Aleatoric uncertainty (AU) is the irreducible uncertainty about the class label remaining after conditioning on the predictive distribution and is given by the conditional entropy $H(Y|\Theta)$,

$$\mathrm{AU}(Q) = H(Y|\Theta) = \mathbb{E}_Q[H(\theta)] = -\int_{\Delta_C} \sum_{c=1}^{C} \theta_c \log \theta_c \, dQ(\theta). \tag{2}$$

Epistemic uncertainty (EU) quantifies uncertainty about which predictive distribution should be assigned to the node and is given by the mutual information (MI) I between Y and Θ,

$$\mathrm{EU}(Q) = I(Y;\Theta) = \mathrm{TU}(Q) - \mathrm{AU}(Q) = \mathbb{E}_Q\left[D_{\mathrm{KL}}(\theta\|\bar{\theta})\right], \tag{3}$$

where $D_{\mathrm{KL}}(\cdot\|\cdot)$ denotes the Kullback-Leibler divergence. We next review graph EDL methods, which directly parameterize Q and are the main second-order baselines considered in this work.

2.1 Graph EDL Models

Evidential node classifiers predict a second-order Dirichlet distribution $Q_\psi^{(i)} = \mathrm{Dir}(\theta; \alpha^{(i)})$, with $\alpha^{(i)} = \mathbf{1} + e^{(i)}$ and class-wise evidence $e^{(i)} \in \mathbb{R}^C_{>0}$. Its mean $\bar{\theta}^{(i)} = \alpha^{(i)}/\alpha_0^{(i)}$ gives the predictive probabilities, while $\alpha_0^{(i)} = \sum_{c=1}^{C} \alpha_c^{(i)}$ controls concentration. We organize graph EDL methods by their evidence parameterization and treatment of epistemic uncertainty: reduction through *feature support* and *structural support*, and reduction or irreducibility due to *neighborhood conflict*.

Direct parameterization. S-BGCN-K (Zhao et al., 2020) directly predicts Dirichlet evidence with a Bayesian GNN and uses dropout-induced variability for uncertainty estimation. During training, its prediction is regularized toward a node-specific, graph-dependent GKDE prior assigning higher evidence to nodes closer to labeled nodes of the corresponding class, thereby encoding structural support; S-BGCN-T-K additionally introduces a teacher network to improve predictive performance. Assuming the Gaussian likelihood and denoting the one-hot ground-truth label $e_y^{(i)}$, its objective is

$$\ell_{\mathrm{MSE}}(\psi; i) = \mathbb{E}_{\theta^{(i)} \sim Q_\psi^{(i)}}\left[\|\theta^{(i)} - e_y^{(i)}\|^2\right] + \lambda D_{\mathrm{KL}}(Q_\psi^{(i)} \| Q_{\mathrm{GKDE}}^{\mathrm{prior},(i)}). \tag{4}$$

Density parameterization. GPN (Stadler et al., 2021) constructs evidence pseudo-counts from class-conditional densities in latent space, so that low-density regions receive little evidence, implementing feature support (Charpentier et al., 2020; 2022). These pseudo-counts are propagated through Personalized PageRank, $\beta_c^{\mathrm{agg},(i)} = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} \beta_c^{\mathrm{ft},(j)}$, yielding $\alpha^{(i)} = \mathbf{1} + \beta^{\mathrm{agg},(i)}$, where $\mathbf{1}$ is a uniform Dirichlet prior. Propagation implements structural support by diffusing evidence across neighboring nodes, while accumulating conflicting class evidence within the same Dirichlet implements neighborhood conflict, attributing disagreement to irreducible aleatoric uncertainty. Assuming the categorical likelihood $p(y^{(i)}|\theta^{(i)}) = \theta_{y^{(i)}}$, GPN minimizes

$$\ell_{\mathrm{UCE}}(\psi; i) = \mathbb{E}_{\theta^{(i)} \sim Q_\psi^{(i)}}[-e_y^{(i)\top} \log \theta^{(i)}] - \lambda H[Q_\psi^{(i)}]. \tag{5}$$

Distance-regularized GPN-GD (Hart et al., 2023) augments this formulation with pairwise regularization of latent representations and evidence.

CUQ-GNN (Damke & Hüllermeier, 2024b) retains the density-based parameterization and UCE objective but replaces the independent encoder and subsequent propagation step with a GNN encoder, jointly learning feature support, structural support, and the handling of neighborhood conflict. GPN-LOP (Damke & Hüllermeier, 2024a) instead retains GPN's feature pseudo-counts but aggregates the resulting Dirichlet distributions, $Q_\psi^{(i)} = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} Q_\psi^{\mathrm{ft},(j)}$, allowing neighborhood disagreement to contribute to epistemic, rather than necessarily aleatoric, uncertainty. It is trained using a tractable upper bound on the UCE loss of the mixture,

$$\ell_{\mathrm{GPN-LOP}}(\psi; i) = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} \left(\mathbb{E}_{\theta^{(j)} \sim Q_\psi^{\mathrm{ft},(j)}}[-e_y^{(i)\top} \log \theta^{(j)}] - \lambda H[Q_\psi^{\mathrm{ft},(j)}]\right). \tag{6}$$

2.2 Why Evaluate Epistemic Uncertainty under Information Growth?

Epistemic uncertainty is commonly considered useful a priori for downstream tasks such as OOD detection on graphs, as the lack of model knowledge about OOD inputs is expected to result in higher epistemic uncertainty (Bazhenov et al., 2023; Wang et al., 2024), and performance on these benchmarks is therefore used to assess the quality of EDL uncertainty estimates. However, our results show that the most effective uncertainty type depends on the particular model and OOD protocol, with AU/TU typically performing best for leave-out-classes (LOC) and feature isolation (FI) with Bernoulli perturbations, and EU for Gaussian FI (Table 1; additional results in Appendix F). That non-epistemic components can outperform EU supports growing criticism that OOD discrimination and the reducibility of epistemic uncertainty under increasing information constitute distinct objectives whose optimal predictors need not coincide (Li et al., 2025; Paplhám et al., 2026). This motivates evaluating epistemic uncertainty under information growth.

Table 1: AUROC for OOD detection on PubMed under LOC and FI protocols. Best-performing uncertainty type(s) for each EDL model and OOD protocol are shown in bold.

	LOC			FI (Bernoulli)			FI (Gaussian)		
Model	TU	AU	EU	TU	AU	EU	TU	AU	EU
CUQ-APPNP	**0.693**	**0.693**	0.643	**0.652**	**0.652**	0.616	0.975	0.942	**0.994**
CUQ-GAT	**0.709**	**0.709**	0.618	**0.586**	**0.586**	0.540	0.660	0.659	**0.684**
CUQ-GCN	**0.717**	**0.717**	0.591	**0.613**	**0.613**	0.554	0.936	0.858	**0.945**
GPN	0.730	0.730	**0.769**	0.593	0.593	**0.638**	0.560	0.560	**0.642**
GPN-GD$_{\text{EVI}}$	0.800	0.771	**0.857**	0.585	0.579	**0.615**	0.578	0.569	**0.621**
GPN-GD$_{\text{LAT}}$	0.724	0.724	**0.764**	0.596	0.596	**0.640**	0.559	0.559	**0.658**
GPN-LOP	0.748	**0.781**	0.651	0.578	0.446	**0.689**	0.696	0.588	**0.767**
S-BGCN-K	0.625	**0.627**	0.500	0.646	**0.679**	0.288	0.342	0.001	**1.000**
S-BGCN-T-K	0.607	**0.611**	0.507	0.654	**0.677**	0.308	0.447	0.006	**1.000**

3 A Statistical Framework for Graph Epistemic Uncertainty

Addressing **(RQ1)**, we extend the i.i.d. characterization of epistemic predictors (Bengs et al., 2022; Juergens et al., 2024) to node classification by defining a statistical framework with an information-growth experimental protocol and a consistency criterion for epistemic predictors.

3.1 Information-Growth Experimental Protocol

Constructing an information-growth experimental protocol for node classification requires three components: (i) the *sampling unit*, i.e., what constitutes one observation from the DGP; (ii) the *sampling scheme*, i.e., how increasingly informative observations are generated from the same DGP; and (iii) the *learning task*, i.e., how the supervised prediction problem evolves along this sequence.

Sampling unit. Unlike classical datasets, the nodes of a graph are not a collection of i.i.d. observations. We therefore take the observed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, equivalently the subgraph induced by $\mathcal{V}$, as the sampling unit, following the graph-as-observation perspective of Shalizi & Rinaldo (2013). Increasing information then corresponds to observing nested graphs $\mathcal{G} \subset \mathcal{G}'$, with $\mathcal{V} \subset \mathcal{V}'$ and $\mathcal{G}$ equal to the subgraph of $\mathcal{G}'$ induced by $\mathcal{V}$.

Sampling scheme. Requiring two nested graphs to arise from the *same* DGP is non-trivial. This consistency is implicit in the i.i.d. setting, where every finite sample is a restriction of the same infinite sequence, but need not hold across graphs of different sizes. We make it explicit through *projectibility* (Shalizi & Rinaldo, 2013). Let $\mathcal{A}$ be a collection of finite subsets of a countable node index set $\mathbb{I}$, ordered by inclusion, and let $\{\mathbb{P}_{A,\phi}\}_{A\in\mathcal{A},\phi\in\Phi}$ denote a family of finite-dimensional DGPs, where $G_A \sim \mathbb{P}_{A,\phi}$ denotes the attributed graph indexed by A. For $A \subset B$, let $\pi_{B\to A}$ denote the natural projection that restricts an observation indexed by B to the indices in A.

Definition 3.1 (Projective DGP)**.** *The family* $\{\mathbb{P}_{A,\phi}\}_{A\in\mathcal{A},\phi\in\Phi}$ *is* projective *if, for every* $A \subset B$*,*

$$\mathbb{P}_{A,\phi} = \mathbb{P}_{B,\phi} \circ \pi_{B\to A}^{-1} \qquad \forall \phi \in \Phi. \tag{7}$$

Thus, projectibility gives a coherent meaning to information growth by ensuring that every smaller observation has exactly the distribution obtained by marginalizing any larger observation under

the stochastic process with the same parameter ϕ. Moreover, by the Kolmogorov extension theorem (Kallenberg, 1997), every such projective family uniquely defines an infinite-dimensional distribution $\mathbb{P}_\phi$ whose finite-dimensional marginals are precisely $\{\mathbb{P}_{A,\phi}\}_{A\in\mathcal{A}}$. Dense graphons are an important class of projective graph families (Veitch & Roy, 2015) and will serve as the concrete instantiation of this framework in our experiments. Other prominent graph models, including many exponential random graph models and sparse random graph formulations, are not projective (Shalizi & Rinaldo, 2013; Schweinberger et al., 2017).

Learning task. Finally, the supervised tasks induced by increasingly large graph observations must exploit the corresponding information growth. If $\mathcal{T}^{(\mathrm{tr})} \subset \mathcal{V}$ and $\mathcal{T}'^{(\mathrm{tr})} \subset \mathcal{V}'$ denote the training sets associated with $\mathcal{G} \subset \mathcal{G}'$, respectively, we require $\mathcal{T}^{(\mathrm{tr})} \subset \mathcal{T}'^{(\mathrm{tr})}$, so that each larger task extends the previous one with additional information about the same underlying node-classification problem. Validation and test sets are defined analogously.

3.2 Consistent Epistemic Predictors

We now characterize a consistent epistemic predictor under this information-growth experimental protocol. For an increasing sequence $A_n \subset A_{n+1}$, with $|A_n| = n$, let $Q_{A_n} = \{Q^{(i)}_{A_n}\}_{i\in A_n}$ denote the collection of node-wise second-order predictors. For an epistemic uncertainty measure U, we define the uncertainty of this collection as the average node-wise uncertainty $\bar{U}_{A_n} := |A_n|^{-1}\sum_{i\in A_n} U(Q^{(i)}_{A_n})$. We define consistency over the full node set $\mathcal{A}_n$, even though empirically we evaluate uncertainty on the test nodes. Let $p^\star_\phi$ denote the Bayes-optimal predictor under the infinite-dimensional law $\mathbb{P}_\phi$ induced by projectibility, and $p^\star_\phi|A_n$ its restriction to A_n.

Definition 3.2 (Consistent epistemic predictors)**.** *Under the information-growth experimental protocol defined in Section 3.1, a sequence of collections of second-order predictors $\{Q_{A_n}\}_{n\geq 1}$ is* consistent *if the following two conditions hold:*

1. ***Monotonicity.*** *For every n, $\mathbb{E}_{G_{A_n}}[\bar{U}_{A_n}] \geq \mathbb{E}_{G_{A_{n+1}}}[\bar{U}_{A_{n+1}}]$. Furthermore, there exist n_0 and $k > 0$ such that $\mathbb{E}_{G_{A_{n_0}}}[\bar{U}_{A_{n_0}}] > \mathbb{E}_{G_{A_{n_0+k}}}[\bar{U}_{A_{n_0+k}}]$.*
2. ***Consistency.*** *As $|A_n| \to \infty$, Q_{A_n} converges in probability to the Dirac measure at $p^\star_\phi|A_n$ under the distance $d(Q_{A_n}, \delta_{p^\star_{\phi|A_n}}) = |A_n|^{-1}\sum_{i\in A_n} d(Q^{(i)}_{A_n}, \delta^{(i)}_{p^\star_{\phi|A_n}})$, where the node-wise distance d metrizes weak convergence on $\mathcal{P}(\Delta_C)$.*

These requirements characterize epistemic uncertainty as reducible with information: by condition (1), it decreases in expectation, excluding constant uncertainty sequences (Juergens et al., 2024), and by (2), it vanishes through concentration on the limiting Bayes-optimal predictor (Bengs et al., 2022). For the entropy-based uncertainty considered here, this yields $\lim_{n\to\infty}\mathbb{E}_{G_{A_n}}[\bar{U}_{A_n})] = 0$, leaving only true aleatoric uncertainty. Such consistency requires strong assumptions (Jain & Bates, 2025); in practice, model misspecification and estimation bias may affect the predictor's estimates without being explicitly represented by its epistemic uncertainty (Jimenez et al., 2026).

4 Graph EDL under Information Growth

We examine graph EDL methods considered in Section 2.1 when estimation and approximation errors are removed. By analyzing their population optima, we study whether they are consistent epistemic predictors per Definition 3.2. Unlike their i.i.d. counterparts, these methods incorporate graph-specific uncertainty principles and losses, motivating an extension of the analogous population-level analysis by Shen et al. (2024). To analyze them within a common framework, we express their objectives through the following unified loss.

Definition 4.1 (Unified EDL loss)**.** *Let Q^{prior} be a fixed Dirichlet prior, $p(y|\theta)$ a likelihood, and $\nu > 0$ a tempering parameter. For each node i, the unified EDL loss is*

$$\ell_{\mathrm{U}}(\psi; i) := D_{\mathrm{KL}}(Q^{(i)}_\psi \| Q_\nu(\cdot|y^{(i)})), \tag{8}$$

where $Q_\nu(\theta|y) \propto Q^{\mathrm{prior}}(\theta)p(y|\theta)^\nu$ denotes the tempered posterior.

The next lemma establishes the equivalence between ℓ_{U} and the training objectives of each method.

Lemma 4.2. *The unified EDL loss ℓ_{U} is equivalent, up to positive scaling and a ψ-independent additive constant, to the EDL objectives under the following choices:*

1. *$p(e_y^{(i)}|\theta^{(i)}) = \mathcal{N}(e_y^{(i)}; \theta^{(i)}, \sigma^2 I_C)$, Q^{prior} is a fixed Dirichlet prior, and $\lambda = 2\sigma^2/\nu$, yielding ℓ_{MSE}.*
2. *$p(y^{(i)}|\theta^{(i)}) = \theta^{(i)}_{y^{(i)}}$, Q^{prior} is a uniform Dirichlet prior, and $\lambda = \nu^{-1}$, yielding ℓ_{UCE}.*
3. *With a uniform Dirichlet prior Q^{prior} and $\lambda = \nu^{-1}$, evaluating the UCE loss at the PPR-weighted soft target $\tilde{e}_y^{(j)}$ yields the reformulated GPN-LOP node-wise objective $\tilde{\ell}_{\mathrm{GPN-LOP}}(\psi, j) \propto \tilde{\ell}_{\mathrm{UCE}}(\psi, j)$; hence, it is equivalent to the unified objective $\tilde{\ell}_{\mathrm{U}}(\psi; j)$.*

We now characterize the population optimum of this common loss.

Theorem 4.3. *The population minimizer of the unified EDL loss ℓ_{U} under $\mathbb{P}_\phi$ satisfies*

$$\arg\min_{\psi} \mathbb{E}_{\mathbb{P}_\phi(\mathbf{A},\mathbf{X},Y)}[\ell_{\mathrm{U}}(\psi; i)] = \arg\min_{\psi} \mathbb{E}_{\mathbb{P}_\phi(\mathbf{A},\mathbf{X})}[D_{\mathrm{KL}}(Q_\psi^{(i)} \| Q^{\star,(i)})], \tag{9}$$

where

$$Q^{\star,(i)}(\theta) = \frac{Q^{\mathrm{prior}}(\theta) \exp(\nu \mathbb{E}_{\mathbb{P}_\phi(y^{(i)}|\mathbf{X},\mathbf{A})}[\log p(y^{(i)}|\theta^{(i)})])}{\int_{\Delta_C} Q^{\mathrm{prior}}(\theta') \exp(\nu \mathbb{E}_{\mathbb{P}_\phi(y^{(i)}|\mathbf{X},\mathbf{A})}[\log p(y^{(i)}|\theta')])\, d\theta'}. \tag{10}$$

The proofs of Lemma 4.2 and Theorem 4.3 are provided in Appendix A.1. For fixed Q^{prior} and finite ν, GPN, GPN-LOP, and CUQ-GNN retain non-vanishing epistemic uncertainty at the population optimum, failing Definition 3.2, while their aleatoric uncertainty depends on ν and may differ from that of the Bayes-optimal predictor; see Appendix A.2.1 for more details on GPN-LOP. For S-BGCN-K, an unbounded increase in local evidence, such as under the dense graphon setting, implies convergence of the population-level GKDE prior and the corresponding population optimum to a Dirac distribution; see Appendix A.2.2. With Monte Carlo dropout, however, epistemic uncertainty need not vanish (Verdoja & Kyrki, 2020). The location of the limiting Dirac also depends on the GKDE construction and need not be Bayes-optimal. Moreover, because they are deterministic models, none of these methods account for epistemic uncertainty induced by data sampling (Shen et al., 2024), which is a recognized source of epistemic uncertainty (Jimenez et al., 2026) and, in our setting, corresponds to randomness from observing a single finite graph. S-BGCN-T-K and GPN-GD additionally regularize class-probability estimation and pairwise relations between training nodes, respectively, without introducing further mechanisms for evidence accumulation under graph growth, and are therefore not analyzed separately.

Thus, the asymptotic behavior of epistemic and aleatoric uncertainty predicted by S-BGCN-K, GPN, GPN-LOP, and CUQ-GNN is determined externally, either by the loss regularization or by the construction of the prior. Graph-specific uncertainty principles defined on a fixed graph therefore do not by themselves guarantee the required information-growth behavior.

5 Graph Bootstrap Ensembles

The previous analysis motivates constructing a second-order predictor whose uncertainty is tied explicitly to the information available about the graph DGP. We pursue this idea through *graph bootstrap ensembles* (GB-Ens), extending deep ensembles, an already understudied approach in node classification (Vieira et al., 2026), to the largely unexplored setting of graph bootstrapping. For each of B bootstrap graph realizations, we train S independently initialized GNNs, mirroring i.i.d. bootstrap ensembles that combine procedural and data variability through resampling (Juergens et al., 2024; Shen et al., 2024; Jain & Bates, 2025; Jimenez et al., 2026). The ability to represent both recognized sources of epistemic uncertainty (Jain & Bates, 2025; Jimenez et al., 2026) provides a key motivation for using bootstrap ensembles rather than alternative second-order predictors, including EDL models and deep ensembles, which may fail to capture one or both sources.

In the i.i.d. setting, bootstrap-based MI has also been shown, under suitable conditions, to converge asymptotically to its Bayesian counterpart (Jain & Bates, 2025), providing theoretical support for

bootstrap-based epistemic uncertainty estimation with neural networks. For graphs, however, bootstrap guarantees remain restricted to specific combinations of graph DGPs, bootstrap schemes, and estimation targets (Schmid & Desmarais, 2017; Levin & Levina, 2019; Green & Shalizi, 2022; Lin et al., 2020; Shao & Le, 2024), with no analogous result for MI. Nevertheless, projective synthetic graph DGPs whose sampling process can be accurately approximated by a bootstrap procedure provide a tractable setting for studying whether graph bootstrap ensembles empirically exhibit the consistent information-growth behavior of Definition 3.2, analogous to that established in the i.i.d. setting. GB-Ens is also computationally feasible despite increased total cost, through parallel training and sequential inference; see Appendix G for computational comparisons.

Uncertainty decomposition. For node i, let $\theta_{b,s}^{(i)} \in \Delta_C$ be the predictive distribution from seed s on bootstrap realization b. The resulting second-order predictor and predictive mean are

$$Q_{B,S}^{(i)} = \frac{1}{BS}\sum_{b=1}^{B}\sum_{s=1}^{S}\delta_{\theta_{b,s}^{(i)}}, \qquad \bar{\theta}^{(i)} = \frac{1}{BS}\sum_{b=1}^{B}\sum_{s=1}^{S}\theta_{b,s}^{(i)}, \tag{11}$$

where $\delta_{\theta_{b,s}^{(i)}}$ denotes the Dirac measure concentrated at $\theta_{b,s}^{(i)}$, with

$$\mathrm{TU}^{(i)} = H(\bar{\theta}^{(i)}), \qquad \mathrm{AU}^{(i)} = \frac{1}{BS}\sum_{b=1}^{B}\sum_{s=1}^{S}H(\theta_{b,s}^{(i)}), \qquad \mathrm{EU}^{(i)} = \mathrm{TU}^{(i)} - \mathrm{AU}^{(i)}. \tag{12}$$

The crossed $B \times S$ design further allows us to separate the two sources of epistemic uncertainty represented by GB-Ens. With the seed-averaged prediction $\bar{\theta}_b^{(i)} = S^{-1}\sum_{s=1}^{S}\theta_{b,s}^{(i)}$,

$$\mathrm{EU}_{\mathrm{boot}}^{(i)} = H(\bar{\theta}^{(i)}) - \frac{1}{B}\sum_{b=1}^{B}H(\bar{\theta}_b^{(i)}), \qquad \mathrm{EU}_{\mathrm{seed}}^{(i)} = \frac{1}{B}\sum_{b=1}^{B}\left[H(\bar{\theta}_b^{(i)}) - \frac{1}{S}\sum_{s=1}^{S}H(\theta_{b,s}^{(i)})\right]. \tag{13}$$

Here, $\mathrm{EU}_{\mathrm{boot}}$ measures disagreement across bootstrap graph realizations, while $\mathrm{EU}_{\mathrm{seed}}$ measures training-induced disagreement within each realization, and $\mathrm{EU}^{(i)} = \mathrm{EU}_{\mathrm{boot}}^{(i)} + \mathrm{EU}_{\mathrm{seed}}^{(i)}$.

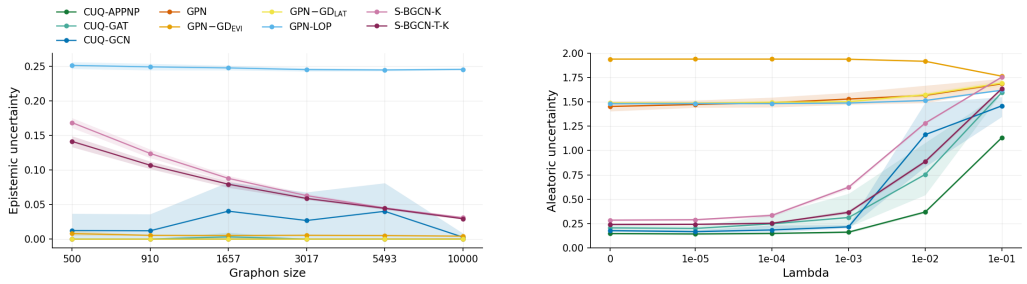
Graph resampling. To instantiate the graph-resampling component of GB-Ens, we adopt a local non-parametric bootstrap adapted from Jeong & Donnat (2025), developed under an exchangeable graphon model. It estimates asymptotically unbiased edge probabilities from the observed connectivity of similar nodes and samples graph realizations accordingly. We use an edge-only bootstrap that retains node identities, features, and labels across realizations, thereby preserving the node classification task while capturing sampling variability in the graph structure. The complete resampling procedure and implementation details are given in Appendix C.

6 Experiments

We empirically address **(RQ2)**–**(RQ3)** through controlled experiments on synthetic projective graph DGPs. For **(RQ2)**, we validate our analysis of graph EDL uncertainty under information growth. For **(RQ3)**, we assess graph bootstrap ensembles as an alternative approach to epistemic prediction.

Models. We evaluate S-BGCN-K and S-BGCN-T-K (Zhao et al., 2020), GPN (Stadler et al., 2021), GPN-GD$_{\mathrm{EVI}}$ and GPN-GD$_{\mathrm{LAT}}$ (Hart et al., 2023), GPN-LOP (Damke & Hüllermeier, 2024a), CUQ-APPNP, CUQ-GAT, and CUQ-GCN (Damke & Hüllermeier, 2024b) as EDL baselines, and GB-Ens with APPNP, GAT, and GCN architectures. Each GB-Ens comprises 25×4 models, corresponding to 25 bootstrap samples and 4 random seeds per sample. Model parameterizations are reported in Appendix D.

Datasets. Our experiments comprise two synthetic projective graphon DGPs. The synthetic DGPs are an exponential-decay graphon and an assortative stochastic-block model (SBM), which define complementary continuous- and discrete-structure regimes for a balanced seven-class node-classification problem. For each DGP, we generate 20 independent replicates of a projective sequence, with six graph sizes logarithmically spaced between 500 and 10,000 nodes. At each graph size, we use $5\%/15\%/80\%$ class-stratified train/validation/test splits constructed according to the protocol in Section 3.1 and fixed across bootstrap replicates for each observed graph. Full DGP specifications and graphon statistics are reported in Appendices B and C.



(a) Epistemic uncertainty under information growth. (b) Aleatoric uncertainty across λ at $N = 10{,}000$.

Figure 1: Uncertainty estimates of EDL models on the exponential-decay graphon.

RQ2: Graph EDL models are inconsistent epistemic predictors.

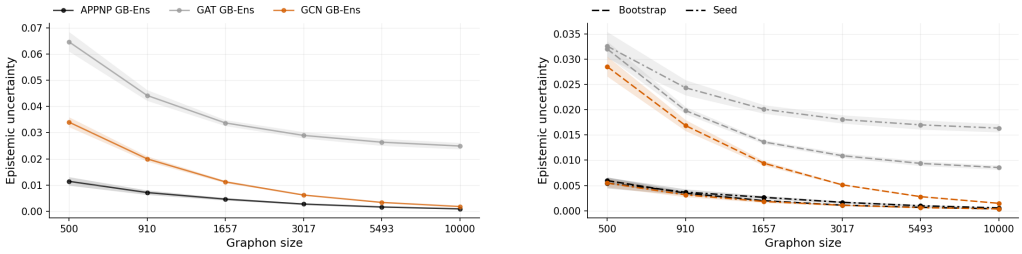
We examine whether the uncertainty behavior of graph EDL models characterized at the nonparametric population optimum by Theorem 4.3 persists with finite data and parametric models. We assess both EU concentration with increasing information and the dependence of EU and AU on model parameterization. Figure 1a shows mean EU across 20 projective-sequence replicates of the exponential-decay graphon, with 95% bootstrap confidence intervals (CIs). S-BGCN-K and S-BGCN-T-K exhibit decreasing EU, as permitted by their information-dependent prior, but remain non-negligible over the observed sequence, whereas GPN, GPN-GD, GPN-LOP, and CUQ-GNN remain approximately constant, consistent with their population optima retaining finite evidence. EU magnitude also differs substantially across methods, reflecting its theoretical dependence on the parameterization, as illustrated by the similar magnitudes of GPN, GPN-GD, and CUQ-GNN under shared parameter values. At $N = 10{,}000$, varying the EDL regularization parameter λ reveals substantial changes in AU on the same observed graph (Figure 1b), demonstrating its dependence on model parameterization rather than on the DGP alone. Analogous results for the SBM are reported in Appendix E.

RQ3: GB-Ens is a family of candidate consistent epistemic predictors.

Figure 2a shows mean GB-Ens epistemic uncertainty across 20 projective-sequence replicates of the exponential-decay graphon, with pointwise 95% bootstrap CIs. From $N = 500$ to $N = 10{,}000$, EU decreases monotonically by 94% for GCN, 91% for APPNP, and 61% for GAT. On the SBM (Appendix E), EU similarly decreases substantially for APPNP and GCN, whereas GAT shows a weaker, non-monotonic reduction due to persistent randomized-training variability, plausibly reflecting its sensitivity to learned attention weights and additional attention dropout. The decomposition in Figure 2b further shows comparable randomized-seed and bootstrap contributions to EU that decrease with increasing information, with relative magnitude depending on the backbone and DGP. Finally, Figure 3b shows that GB-Ens AU decreases and eventually stabilizes, with all three backbones approaching similar plateaus, consistent with convergence toward a predictor-independent estimate of aleatoric uncertainty. The same behavior is observed for the SBM (Appendix E).

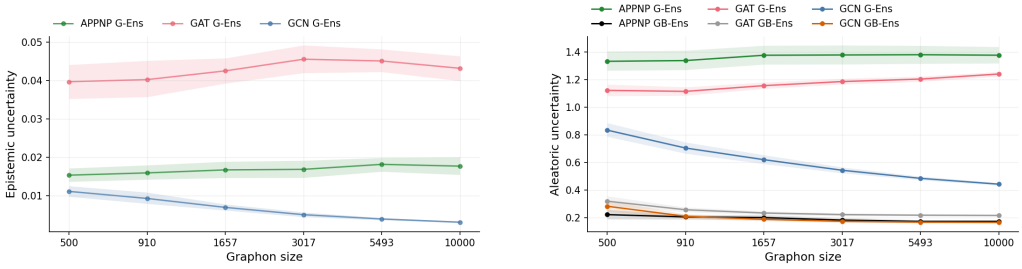
To assess the effect of graph resampling, we compare GB-Ens with standard graph deep ensembles (G-Ens), trained with 10 different random seeds on the observed graph using the same backbones. Figure 3a shows that, on the exponential-decay graphon, G-Ens exhibit no consistent EU contraction, while Figure 3b shows strongly backbone-dependent AU trajectories, with GAT trending upward and GCN continuing to decrease. Similar behavior is observed on the SBM (Appendix E). Moreover, the gap between G-Ens EU and the randomized-training component of GB-Ens widens with graph size, as the latter contracts under information growth while the former does not. Graph resampling therefore does not simply add bootstrap variability, but induces convergent behavior also in the seed component, enabling EU contraction that is not observed for standard deep ensembles.

As a simple diagnostic of GB-Ens's approximation to the full second-order distribution induced by true-DGP resampling, which captures predictive variability without DGP estimation error, we compare the two distributions using normalized node-wise Wasserstein-1 distances. On the exponential-



(a) Total epistemic uncertainty. (b) Bootstrap and training-seed decomposition.

Figure 2: GB-Ens mean epistemic uncertainty on the exponential-decay graphon.



(a) Epistemic uncertainty of G-Ens. (b) Aleatoric uncertainty of GB-Ens and G-Ens.

Figure 3: Ensemble mean uncertainties on the exponential-decay graphon.

decay graphon, distances remain small and stable with graph size for APPNP (≈ 0.29) and GCN (≈ 0.21), indicating good agreement, while increasing from 0.45 to 0.62 for GAT. Similar results on the SBM are reported in Appendix E, suggesting that approximation quality varies across predictors.

Overall, these results support GB-Ens as a family of candidate consistent epistemic predictors. This behavior nevertheless remains predictor-specific, as EU contraction ultimately relies on regularity of the underlying learning procedure; persistent training variability can therefore produce weaker and partly non-monotonic reductions.

7 Conclusions

Take-home message. Epistemic uncertainty in node classification should be assessed through its behavior under increasing information, rather than solely by obeying principles or being useful for downstream tasks on a fixed graph. This perspective identifies consistent epistemic uncertainty as requiring estimates that reflect data uncertainty rather than remaining hyperparameter-dependent, a requirement generally not met by graph EDL models. GB-Ens provides a constructive approach by capturing data and procedural uncertainty, with controlled experiments supporting it as a family of candidate consistent predictors.

Limitations & future work. The theoretical consistency of bootstrap-based graph epistemic predictors remains to be established. Moreover, information-growth behavior cannot be evaluated on standard real-world benchmarks, which lack the sampling process required to generate increasingly informative observations from the same DGP. Benchmarks leveraging real graph-growth mechanisms, where projectibility is often natural (Shalizi & Rinaldo, 2013), could extend our analysis beyond synthetic random graphs, either retaining node classification and motivating improved bootstrap-based predictors or considering tasks more naturally aligned with graph growth.

AI USE STATEMENT

In this work, we used generative AI tools for research ideation and to assist in proving mathematical claims. We did not use generative AI tools for generating synthetic datasets, and the remaining required-disclosure tasks were not applicable to this work. Additionally, we used generative AI tools to aid and polish the writing and for retrieval and discovery, including identifying related work. All AI-assisted suggestions and outputs were reviewed by the authors for correctness and appropriateness before being incorporated into the work. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The authors are not aware of any specific ethical concerns raised by this work. The study does not involve human subjects or the collection of sensitive data, and the experiments are primarily conducted on synthetic graph data and established benchmark datasets.

REPRODUCIBILITY STATEMENT

An anonymized implementation, configuration files, and scripts for reproducing the reported experiments are available at https://anonymous.4open.science/r/F0B4 and in the supplementary material. Section 3 defines the statistical framework for epistemic uncertainty under information growth; Section 3.1 specifies the information-growth experimental protocol and Definition 3.1 formalizes projective graph data-generating processes, while Definition 3.2 defines consistent epistemic predictors. Section 4 gives the unified analysis of graph EDL methods and their population optima, with complete proofs and model-specific derivations provided in Appendix A. Section 5 defines graph bootstrap ensembles, their uncertainty decomposition, and the graph-resampling construction, whose complete algorithmic specification is given in Appendix C. Section 6 reports the experimental protocol, evaluated models, graph sizes, train/validation/test splits, and number of independent projective-sequence replicates. Appendix B provides the complete synthetic graph-DGP specifications, generator parameters, and realized graph statistics, while Appendix D documents model architectures, optimization settings, and hyperparameters. Additional information-growth results are reported in Appendix E, the OOD evaluation protocols, datasets, and metrics in Appendix F, and computational-complexity and empirical runtime/memory analyses in Appendix G.

REFERENCES

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

Gleb Bazhenov, Denis Kuznedelev, Andrey Malinin, Artem Babenko, and Liudmila Prokhorenkova. Revisiting uncertainty estimation for node classification: New benchmark and insights, 2023. URL https://openreview.net/forum?id=DB3BH3arU2Y.

Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Pitfalls of epistemic uncertainty quantification through loss minimisation. *Advances in Neural Information Processing Systems*, 35: 29205–29216, 2022.

Bertrand Charpentier, Daniel Zügner, and Stephan Günnemann. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. *Advances in neural information processing systems*, 33:1356–1367, 2020.

Bertrand Charpentier, Oliver Borchert, Daniel Zügner, Simon Geisler, and Stephan Günnemann. Natural posterior network: Deep bayesian predictive uncertainty for exponential family distributions. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=tV3N0DWMxCg.

Chao Chen, Chenghua Guo, Rui Xu, Jiujiu Chen, Xiangwen Liao, Xi Zhang, Sihong Xie, Hui Xiong, and Philip Yu. Uncertainty quantification on graph learning: a survey. *arXiv preprint arXiv:2404.14642*, 2024.

Clemens Damke and Eyke Hüllermeier. Linear opinion pooling for uncertainty quantification on graphs. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024a. URL https://openreview.net/forum?id=qLGkfpXTSn.

Clemens Damke and Eyke Hüllermeier. Cuq-gnn: Committee-based graph uncertainty quantification using posterior networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 306–323. Springer, 2024b.

Stefan Depeweg. *Modeling epistemic and aleatoric uncertainty with bayesian neural networks and latent variables*. PhD thesis, Technische Universität München, 2019.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Combining neural networks with personalized pagerank for classification on graphs. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=H1gL-2A9Ym.

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks: J. gawlikowski et al. *Artificial intelligence review*, 56(Suppl 1):1513–1589, 2023.

C Lee Giles, Kurt D Bollacker, and Steve Lawrence. Citeseer: An automatic citation indexing system. In *Proceedings of the third ACM conference on Digital libraries*, pp. 89–98, 1998.

Alden Green and Cosma Rohilla Shalizi. Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics*, 16(1):1058–1095, 2022.

Russell Hart, Linlin Yu, Yifei Lou, and Feng Chen. Improvements on uncertainty quantification for node classification via distance based regularization. *Advances in Neural Information Processing Systems*, 36:55454–55478, 2023.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021.

Anchit Jain and Stephen Bates. Deep ensembles for epistemic uncertainty: A frequentist perspective. *arXiv preprint arXiv:2510.22063*, 2025.

Sowon Jeong and Claire Donnat. LOBSTUR: A local bootstrap framework for tuning unsupervised representations in graph neural networks. In *Machine Learning on Graphs in the Era of Generative Artificial Intelligence*, 2025. URL https://openreview.net/forum?id=xRasvIyMcW.

Sebastian Jimenez, Mira Juergens, and Willem Waegeman. Position: Epistemic uncertainty estimation methods are fundamentally incomplete. In *Forty-third International Conference on Machine Learning Position Paper Track*, 2026. URL https://openreview.net/forum?id=g598HZM6ib.

Mira Juergens, Nis Meinert, Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Is epistemic uncertainty faithfully represented by evidential deep learning methods? In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=mxjB0LIgpT.

Olav Kallenberg. *Foundations of modern probability*. Springer, 1997.

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

Keith Levin and Elizaveta Levina. Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*, 2019.

Yucen Lily Li, Daohan Lu, Polina Kirichenko, Shikai Qiu, Tim GJ Rudner, C Bayan Bruss, and Andrew Gordon Wilson. Position: Supervised classifiers answer the wrong questions for ood detection. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025.

Qiaohui Lin, Robert Lunde, and Purnamrita Sarkar. Trading off accuracy for speedup: Multiplier bootstraps for subgraph counts. *arXiv preprint arXiv:2009.06170*, 2020.

Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(2):127–163, 2000.

Galileo Namata, Ben London, Lise Getoor, and Bert Huang. Query-driven active surveying for collective classification. 2012. URL https://api.semanticscholar.org/CorpusID:15384610.

Jakub Paplhám, Willem Waegeman, Eyke Hüllermeier, and Vojtěch Franc. Evaluating epistemic uncertainty: Beyond ood detection and active learning. *arXiv preprint arXiv:2607.14817*, 2026.

Yusuf Sale, Nikita Kotelevskii, Maxim Panov, and Eyke Hüllermeier. The aleatoric-epistemic dichotomy of uncertainty is meaningful and indispensable for machine learning. *Available at SSRN 6783338*, 2026.

Christian S Schmid and Bruce A Desmarais. Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *2017 IEEE international conference on big data (Big Data)*, pp. 116–121. IEEE, 2017.

Michael Schweinberger, Pavel N Krivitsky, and Carter T Butts. A note on the role of projectivity in likelihood-based inference for random graph models. *arXiv preprint arXiv:1707.00211*, 2017.

Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. *Annals of statistics*, 41(2):508, 2013.

Zhixuan Shao and Can M Le. Parametric bootstrap for fixed edge-probability network models. *arXiv preprint arXiv:2402.01866*, 2024.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

Maohao Shen, Jongha Jon Ryu, Soumya Ghosh, Yuheng Bu, Prasanna Sattigeri, Subhro Das, and Gregory W Wornell. Are uncertainty quantification capabilities of evidential deep learning a mirage? *Advances in Neural Information Processing Systems*, 37:107830–107864, 2024.

Maximilian Stadler, Bertrand Charpentier, Simon Geisler, Daniel Zügner, and Stephan Günnemann. Graph posterior network: Bayesian predictive uncertainty for node classification. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 18033–18048. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/95b431e51fc53692913da5263c214162-Paper.pdf.

Victor Veitch and Daniel M Roy. The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*, 2015.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Francesco Verdoja and Ville Kyrki. Notes on the behavior of mc dropout. *arXiv preprint arXiv:2008.02627*, 2020.

Pedro C Vieira, Pedro Ribeiro, and Viacheslav Borovitskiy. Do deep ensembles actually capture uncertainty in graph neural networks? *arXiv preprint arXiv:2605.22593*, 2026.

Fangxin Wang, Yuqing Liu, Kay Liu, Yibo Wang, Sourav Medya, and Philip S. Yu. Uncertainty in graph neural networks: A survey. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=0e1Kn76HM1.

Kaizheng Wang, Fabio Cuzzolin, Keivan Shariatmadar, David Moens, and Hans Hallez. A review of uncertainty representation and quantification in neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 2025.

Lisa Wimmer, Yusuf Sale, Paul Hofman, Bernd Bischl, and Eyke Hüllermeier. Quantifying aleatoric and epistemic uncertainty in machine learning: Are conditional entropy and mutual information appropriate measures? In *Uncertainty in artificial intelligence*, pp. 2282–2292. PMLR, 2023.

Hongteng Xu, Dixin Luo, Lawrence Carin, and Hongyuan Zha. Learning graphons via structured gromov-wasserstein barycenters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 10505–10513, 2021.

Xujiang Zhao, Feng Chen, Shu Hu, and Jin-Hee Cho. Uncertainty aware semi-supervised learning on graph data. *Advances in neural information processing systems*, 33:12827–12836, 2020.

A Graph EDL Under Information Growth

A.1 Unified EDL Loss

In this section, we extend the unified formulation of EDL losses from the i.i.d. setting (Shen et al., 2024) to node classification. We begin by defining a unified EDL loss and establishing its equivalence to the MSE and UCE losses used by existing graph EDL methods.

Let Q^{prior} be a fixed Dirichlet prior and $p(y|\theta)$ a likelihood. For a tempering parameter $\nu > 0$, define the tempered posterior

$$Q_\nu(\theta|y) := \frac{Q^{\text{prior}}(\theta)p(y|\theta)^\nu}{\int_{\Delta_C} Q^{\text{prior}}(\theta')p(y|\theta')^\nu d\theta'}. \tag{1}$$

For each node i, the unified EDL loss is

$$\ell_{\text{U}}(\psi; i) := D_{\text{KL}}(Q_\psi^{(i)} \| Q_\nu(\cdot|y^{(i)})). \tag{2}$$

Expanding the KL divergence gives

$$\ell_{\text{U}}(\psi; i) = D_{\text{KL}}(Q_\psi^{(i)} \| Q^{\text{prior}}) - \nu \mathbb{E}_{\theta^{(i)} \sim Q_\psi^{(i)}}[\log p(y^{(i)}|\theta^{(i)})] + C, \tag{3}$$

where C is independent of ψ.

Lemma A.1. *For the Gaussian likelihood $p(e_y^{(i)}|\theta^{(i)}) = \mathcal{N}(e_y^{(i)}; \theta^{(i)}, \sigma^2 I_C)$ and a fixed Dirichlet prior Q^{prior}, setting $\lambda = 2\sigma^2/\nu$ yields*

$$\ell_{\text{U}}(\psi; i) = \lambda^{-1} \ell_{\text{MSE}}(\psi; i) + C, \tag{4}$$

where

$$\ell_{\text{MSE}}(\psi; i) = \mathbb{E}_{\theta^{(i)} \sim Q_\psi^{(i)}}[\|\theta^{(i)} - e_y^{(i)}\|^2] + \lambda D_{\text{KL}}(Q_\psi^{(i)} \| Q^{\text{prior}}), \tag{5}$$

and C is independent of ψ. Thus, ℓ_{U} is equivalent, up to scaling and additive constants, to ℓ_{MSE}.

Proof. Since

$$-\log p(e_y^{(i)}|\theta^{(i)}) = \frac{1}{2\sigma^2}\|\theta^{(i)} - e_y^{(i)}\|^2 + C', \tag{6}$$

substitution into equation 3 gives the result. □

Lemma A.2. *For the categorical likelihood $p(y^{(i)}|\theta^{(i)}) = \theta_{y^{(i)}}^{(i)}$ and a uniform Dirichlet prior Q^{prior}, setting $\lambda = \nu^{-1}$ yields*

$$\ell_{\text{U}}(\psi; i) = \nu \ell_{\text{UCE}}(\psi; i) + C, \tag{7}$$

where

$$\ell_{\text{UCE}}(\psi; i) = \mathbb{E}_{\theta^{(i)} \sim Q_\psi^{(i)}}[-e_y^{(i)\top} \log \theta^{(i)}] - \lambda H(Q_\psi^{(i)}), \tag{8}$$

and C is independent of ψ. Thus, ℓ_{U} is equivalent, up to scaling and additive constants, to ℓ_{UCE}.

Proof. The result follows directly from equation 3 and $D_{\mathrm{KL}}(Q_\psi^{(i)} \| Q^{\mathrm{prior}}) = -H(Q_\psi^{(i)}) + C'$ for a uniform Dirichlet prior. □

Lemma A.3. *For each node $j \in \mathcal{V}$ with*

$$w_j := \sum_{i \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{i,j}^{\mathrm{PPR}} > 0, \tag{9}$$

define the soft target

$$\tilde{e}_y^{(j)} := \frac{1}{w_j} \sum_{i \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{i,j}^{\mathrm{PPR}} e_y^{(i)}. \tag{10}$$

The reformulated GPN-LOP node-wise loss satisfies

$$\tilde{\ell}_{\mathrm{GPN-LOP}}(\psi; j) = w_j \tilde{\ell}_{\mathrm{UCE}}(\psi; j), \tag{11}$$

where

$$\tilde{\ell}_{\mathrm{UCE}}(\psi; j) := \mathbb{E}_{\theta^{(j)} \sim Q_\psi^{\mathrm{ft},(j)}}[-\tilde{e}_y^{(j)\top} \log \theta^{(j)}] - \lambda H(Q_\psi^{\mathrm{ft},(j)}). \tag{12}$$

Let

$$\tilde{\ell}_{\mathrm{U}}(\psi; j) := D_{\mathrm{KL}}(Q_\psi^{\mathrm{ft},(j)} \| \tilde{Q}_\nu^{(j)}), \tag{13}$$

where $\tilde{Q}_\nu^{(j)}$ denotes the tempered posterior under the soft-target categorical likelihood, $p(\tilde{e}_y^{(j)} | \theta^{(j)}) \propto \prod_{c=1}^{C} (\theta_c^{(j)})^{\tilde{e}_{y,c}^{(j)}}$. Hence, under this soft-target likelihood, a uniform Dirichlet prior Q^{prior}, and $\lambda = \nu^{-1}$,

$$\tilde{\ell}_{\mathrm{U}}(\psi; j) = \frac{\nu}{w_j} \tilde{\ell}_{\mathrm{GPN-LOP}}(\psi; j) + C, \tag{14}$$

where C is independent of ψ. Thus, $\tilde{\ell}_{\mathrm{GPN-LOP}}$ is equivalent, up to scaling and an additive constant, to the unified loss $\tilde{\ell}_{\mathrm{U}}$ associated with the soft target $\tilde{e}_y^{(j)}$.

Proof. Starting from the training objective,

$$\mathcal{L}(\psi) = \sum_{i \in \mathcal{T}^{(\mathrm{tr})}} \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} \left(\mathbb{E}_{\theta^{(j)} \sim Q_\psi^{\mathrm{ft},(j)}}[-e_y^{(i)\top} \log \theta^{(j)}] - \lambda H[Q_\psi^{\mathrm{ft},(j)}] \right), \tag{15}$$

interchanging the finite sums gives

$$\mathcal{L}(\psi) = \sum_{j \in \mathcal{V}} \tilde{\ell}_{\mathrm{GPN-LOP}}(\psi; j). \tag{16}$$

For fixed j, linearity of expectation yields

$$\tilde{\ell}_{\mathrm{GPN-LOP}}(\psi; j) = \mathbb{E}_{\theta^{(j)} \sim Q_\psi^{\mathrm{ft},(j)}} \left[-\Big(\sum_{i \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{i,j}^{\mathrm{PPR}} e_y^{(i)} \Big)^\top \log \theta^{(j)} \right] \tag{17}$$

$$- \lambda \sum_{i \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{i,j}^{\mathrm{PPR}} H(Q_\psi^{\mathrm{ft},(j)}) \tag{18}$$

$$= w_j \left(\mathbb{E}_{\theta^{(j)} \sim Q_\psi^{\mathrm{ft},(j)}}[-\tilde{e}_y^{(j)\top} \log \theta^{(j)}] - \lambda H(Q_\psi^{\mathrm{ft},(j)}) \right) \tag{19}$$

$$= w_j \tilde{\ell}_{\mathrm{UCE}}(\psi; j). \tag{20}$$

Applying Lemma A.2 to this soft-target extension,

$$\tilde{\ell}_{\mathrm{U}}(\psi; j) = \nu \tilde{\ell}_{\mathrm{UCE}}(\psi; j) + C, \tag{21}$$

and substituting $\tilde{\ell}_{\mathrm{UCE}}(\psi; j) = w_j^{-1} \tilde{\ell}_{\mathrm{GPN-LOP}}(\psi; j)$ gives the result. □

We next characterize the unconstrained minimizer of the unified objective.

Theorem A.4. *The population minimizer of the unified EDL loss ℓ_{U} under $\mathbb{P}_\phi$ satisfies*

$$\arg\min_{\psi} \mathbb{E}_{\mathbb{P}_\phi(\mathbf{X},\mathbf{A},Y)}[D_{\mathrm{KL}}(Q_\psi^{(i)} \| Q_\nu(\cdot|y^{(i)}))] = \arg\min_{\psi} \mathbb{E}_{\mathbb{P}_\phi(\mathbf{A},\mathbf{X})}[D_{\mathrm{KL}}(Q_\psi^{(i)} \| Q^{\star,(i)})], \tag{22}$$

where

$$Q^{\star,(i)}(\theta) = \frac{Q^{\mathrm{prior}}(\theta) \exp(\nu \mathbb{E}_{\mathbb{P}_\phi(y^{(i)}|\mathbf{X},\mathbf{A})}[\log p(y^{(i)}|\theta^{(i)})])}{\int_{\Delta_C} Q^{\mathrm{prior}}(\theta') \exp(\nu \mathbb{E}_{\mathbb{P}_\phi(y^{(i)}|\mathbf{X},\mathbf{A})}[\log p(y^{(i)}|\theta')])\, d\theta'}. \tag{23}$$

Proof. The proof follows analogously to the i.i.d. case in Theorem 5.1 of Shen et al. (2024), with $(\mathbf{A}, \mathbf{X})$ replacing the i.i.d. covariates. □

A.2 Model-Specific Population Optima

A.2.1 GPN-LOP

For GPN-LOP, the population-optimal feature-based Dirichlet distribution at node j is determined by the PPR-weighted conditional class probabilities $\tilde{\eta}^{(j)} = w_j^{-1} \sum_{k \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{k,j}^{\mathrm{PPR}} \eta^{(k)}$, where $w_j = \sum_{k \in \mathcal{T}^{(\mathrm{tr})}} \Pi_{k,j}^{\mathrm{PPR}}$ and $\eta^{(k)} = \mathbb{E}_{\mathbb{P}_\phi(y^{(k)}|\mathbf{A},\mathbf{X})}[e_y^{(k)}]$. Under the categorical likelihood and uniform Dirichlet prior, Theorem A.4 therefore gives $Q^{\mathrm{ft}\star,(j)} = \mathrm{Dir}\big(\theta; \alpha^{\mathrm{prior}} + \nu\tilde{\eta}^{(j)}\big)$. Since $\sum_{c=1}^{C} \tilde{\eta}_c^{(j)} = 1$, each component has total concentration $\alpha_0^{\mathrm{prior}} + \nu$, which remains finite for fixed ν. GPN-LOP aggregates these feature-based distributions through PPR, $Q^{\star,(i)} = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} Q^{\mathrm{ft}\star,(j)}$. Its epistemic uncertainty decomposes as

$$\mathrm{EU}(Q^{(i)}) = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} \mathrm{EU}(Q^{\mathrm{ft},(j)}) + \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} D_{\mathrm{KL}}(\bar{\theta}^{(j)} \| \bar{\theta}^{(i)}), \tag{24}$$

where $\bar{\theta}^{(i)} = \sum_{j \in \mathcal{V}} \Pi_{i,j}^{\mathrm{PPR}} \bar{\theta}^{(j)}$.

The first term is strictly positive for finite ν, since each Dirichlet component has finite concentration, while the second term is non-negative. Hence, the GPN-LOP population optimum retains strictly positive epistemic uncertainty even when the component means agree.

A.2.2 S-BGCN-K

Theorem A.4 characterizes the population optimum for a fixed prior and therefore does not directly apply to S-BGCN-K, whose GKDE prior depends on the observed graph and training labels. Marginalizing this data-dependent prior in the population objective yields an effective prior $Q_n^{\mathrm{prior},(i)}$, defined, for the data-dependent GKDE prior $Q_{G_{A_n}}^{\mathrm{prior},(i)}$, by

$$Q_n^{\mathrm{prior},(i)}(\theta) \propto \exp(\mathbb{E}_{G_{A_n} \sim \mathbb{P}_{A_n,\phi}}[\log Q_{G_{A_n}}^{\mathrm{prior},(i)}(\theta)]). \tag{25}$$

This effective prior changes with graph size n. We first show that convergence of these priors implies convergence of the corresponding population optima, and then establish conditions under which the effective GKDE prior concentrates under graph growth.

Lemma A.5 (Population optima under convergent priors)**.** *Let $\{Q_n^{\mathrm{prior},(i)}\}_{n \geq 1}$ be a sequence of priors such that $Q_n^{\mathrm{prior},(i)} \Rightarrow Q_\infty^{\mathrm{prior},(i)}$. For fixed finite ν, suppose that $g^{(i)}$ is continuous and strictly positive on Δ_C. Then, the corresponding population optima*

$$Q_n^{\star,(i)}(d\theta) = \frac{g^{(i)}(\theta) Q_n^{\mathrm{prior},(i)}(d\theta)}{\int_{\Delta_C} g^{(i)}(\vartheta) Q_n^{\mathrm{prior},(i)}(d\vartheta)} \tag{26}$$

satisfy $Q_n^{\star,(i)} \Rightarrow Q_\infty^{\star,(i)}$, where $Q_\infty^{\star,(i)}$ is the population optimum of Theorem A.4 associated with the limiting prior $Q_\infty^{\mathrm{prior},(i)}$, namely,

$$Q_\infty^{\star,(i)}(d\theta) = \frac{g^{(i)}(\theta) Q_\infty^{\mathrm{prior},(i)}(d\theta)}{\int_{\Delta_C} g^{(i)}(\vartheta) Q_\infty^{\mathrm{prior},(i)}(d\vartheta)}. \tag{27}$$

Moreover, $Q_\infty^{\star,(i)}$ is a Dirac distribution if and only if $Q_\infty^{\mathrm{prior},(i)}$ is a Dirac distribution.

Proof. For any bounded continuous $f : \Delta_C \to \mathbb{R}$,

$$\int_{\Delta_C} f(\theta) Q_n^{\star,(i)}(d\theta) = \frac{\int_{\Delta_C} f(\theta) g^{(i)}(\theta) Q_n^{\mathrm{prior},(i)}(d\theta)}{\int_{\Delta_C} g^{(i)}(\theta) Q_n^{\mathrm{prior},(i)}(d\theta)}. \tag{28}$$

Since Δ_C is compact, continuity of $g^{(i)}$ implies that $g^{(i)}$ and $fg^{(i)}$ are bounded and continuous. The assumed weak convergence $Q_n^{\mathrm{prior},(i)} \Rightarrow Q_\infty^{\mathrm{prior},(i)}$ therefore gives

$$\int f g^{(i)} dQ_n^{\mathrm{prior},(i)} \to \int f g^{(i)} dQ_\infty^{\mathrm{prior},(i)} \quad \text{and} \quad \int g^{(i)} dQ_n^{\mathrm{prior},(i)} \to \int g^{(i)} dQ_\infty^{\mathrm{prior},(i)}. \tag{29}$$

Since $g^{(i)}$ is strictly positive on Δ_C, the latter limit is strictly positive, and hence $Q_n^{\star,(i)} \Rightarrow Q_\infty^{\star,(i)}$. Moreover, strict positivity implies that $Q_\infty^{\star,(i)}$ and $Q_\infty^{\mathrm{prior},(i)}$ have the same null sets, and hence one is concentrated at a single point if and only if the other is. □

For the Gaussian likelihood underlying ℓ_{MSE}, we define the likelihood-derived weighting function $g^{(i)}(\theta) = \exp\{\nu \, \mathbb{E}_{\mathbb{P}_\phi(y^{(i)}|\mathbf{X},\mathbf{A})}[\log p(y^{(i)}|\theta)]\}$, which is continuous and strictly positive on the compact simplex. Hence, Lemma A.5 applies to any convergent sequence of priors. We next give sufficient conditions for a sequence of Dirichlet priors to concentrate.

Lemma A.6 (Concentration of Dirichlet priors)**.** *Let*

$$Q_n^{\mathrm{prior},(i)} = \mathrm{Dir}(\theta; \beta_n^{(i)}), \qquad \beta_{0,n}^{(i)} = \sum_{c=1}^{C} \beta_{n,c}^{(i)}, \qquad m_n^{(i)} = \frac{\beta_n^{(i)}}{\beta_{0,n}^{(i)}}. \tag{30}$$

If $\beta_{0,n}^{(i)} \to \infty$ and $m_n^{(i)} \to m^{(i)} \in \Delta_C$, then $Q_n^{\mathrm{prior},(i)} \Rightarrow \delta_{m^{(i)}}$.

Proof. Let $\Theta_n^{(i)} \sim Q_n^{\mathrm{prior},(i)}$. Then, $\mathbb{E}[\Theta_n^{(i)}] = m_n^{(i)}$. Moreover, since $m_{n,c}^{(i)} \in [0,1]$, we have $m_{n,c}^{(i)}(1 - m_{n,c}^{(i)}) \le 1/4$, and thus, for each class c,

$$\mathrm{Var}(\Theta_{n,c}^{(i)}) = \frac{m_{n,c}^{(i)}(1 - m_{n,c}^{(i)})}{\beta_{0,n}^{(i)} + 1} \le \frac{1}{4(\beta_{0,n}^{(i)} + 1)} \to 0, \tag{31}$$

where the convergence follows from $\beta_{0,n}^{(i)} \to \infty$. Therefore, $\Theta_n^{(i)} - m_n^{(i)} \to 0$ in probability. Since $m_n^{(i)} \to m^{(i)}$, it follows that $\Theta_n^{(i)} \to m^{(i)}$ in probability, and hence $Q_n^{\mathrm{prior},(i)} \Rightarrow \delta_{m^{(i)}}$. □

Thus, if $\beta_{0,n}^{(i)} \to \infty$ and $m_n^{(i)} \to m^{(i)}$, Lemmas A.6 and A.5 give $Q_n^{\mathrm{prior},(i)} \Rightarrow \delta_{m^{(i)}}$ and $Q_n^{\star,(i)} \Rightarrow \delta_{m^{(i)}}$. Consequently,

$$\mathrm{EU}(Q_n^{\star,(i)}) \to 0 \quad \text{and} \quad \mathrm{AU}(Q_n^{\star,(i)}) \to H(m^{(i)}). \tag{32}$$

S-BGCN-K provides a graph-dependent Dirichlet prior through GKDE, with parameters

$$\beta_{n,c}^{(i)} = 1 + \sum_{j \in \mathcal{T}_n^{(\mathrm{tr})}} \mathbf{1}\{y^{(j)} = c\} g_\sigma(d_{ij}), \tag{33}$$

where d_{ij} denotes the graph distance between nodes i and j, and $g_\sigma(d_{ij})$ is the Gaussian kernel applied to this distance, with bandwidth σ. Since the expectation of a Dirichlet log-density differs from the log-density of a Dirichlet distribution with expected parameters only by a θ-independent constant, the effective population prior is

$$Q_n^{\mathrm{prior},(i)} = \mathrm{Dir}(\theta; \beta_n^{(i)}), \qquad \beta_{n,c}^{(i)} = \mathbb{E}_{G_{A_n} \sim \mathbb{P}_{A_n,\phi}}[\beta_{n,c}^{(i)}]. \tag{34}$$

Summing over classes gives the total concentration of the effective prior,

$$\beta_{0,n}^{(i)} = C + \mathbb{E}_{G_{A_n} \sim \mathbb{P}_{A_n,\phi}}\Big[\sum_{j \in \mathcal{T}_n^{(\mathrm{tr})}} g_\sigma(d_{ij}) \Big]. \tag{35}$$

Thus, concentration grows with the expected amount of kernel-weighted labeled evidence available around node i. In particular, for any fixed finite D,

$$\beta_{0,n}^{(i)} \geq C + g_\sigma(D)\mathbb{E}_{G_{A_n}\sim\mathbb{P}_{A_n,\phi}}[|\{j \in \mathcal{T}_n^{(\mathrm{tr})} : d_{ij} \leq D\}|], \tag{36}$$

where we use that the Gaussian kernel is non-increasing in distance. Hence, a sufficient condition for $\beta_{0,n}^{(i)} \to \infty$ is the existence of a fixed D such that $\mathbb{E}_{G_{A_n}\sim\mathbb{P}_{A_n,\phi}}[|\{j \in \mathcal{T}_n^{(\mathrm{tr})} : d_{ij} \leq D\}|] \to \infty$. Under the dense projective graph DGPs considered in our experiments, positive population normalized degree for node i, together with a training-label sampling scheme that retains a positive fraction of its neighbors, yields a diverging expected number of labeled one-hop neighbors, so the condition holds with $D = 1$. If, in addition, the normalized GKDE evidence converges, $m_n^{(i)} \to m^{(i)}$, Lemma A.6 gives $Q_n^{\mathrm{prior},(i)} \Rightarrow \delta_{m^{(i)}}$ and Lemma A.5 then yields $Q_n^{\star,(i)} \Rightarrow \delta_{m^{(i)}}$.

However, $m^{(i)}$ is the limiting vector of normalized population-averaged kernel-weighted class evidence induced by the GKDE construction and is not guaranteed to coincide with the Bayes-optimal predictor $p_\phi^{\star,(i)}$, as it depends on choices such as the GKDE bandwidth and the graph-distance weighting function. Hence, despite concentration of the evidential distribution, Definition 3.2 need not be satisfied, and $\mathrm{AU}(Q_n^{\star,(i)}) \to H(m^{(i)})$ need not equal the true aleatoric uncertainty $H(p_\phi^{\star,(i)})$.

The analysis above considers S-BGCN-K with Monte Carlo (MC) dropout disabled, thereby isolating the uncertainty induced by the evidential objective. When MC dropout with a fixed Bernoulli dropout rate is enabled, as in the original formulation, concentration of the evidential distribution is no longer sufficient for concentration of the resulting predictive distribution: the predictions must additionally become invariant to the dropout mask as information grows. Fixed-rate MC dropout does not in general ensure such concentration, and has been shown to retain predictive variability independently of the amount of observed data in simple settings (Verdoja & Kyrki, 2020).

B Synthetic Graphons

We consider two complementary homophilic graph DGPs: a smooth exponential-decay graphon, representing continuous latent similarity, and an assortative stochastic block model (SBM) graphon, representing discrete community structure, both commonly considered in the literature (Xu et al., 2021). Besides providing qualitatively different graph structure, the SBM connects our construction to projective dyadic-independence models studied within the exponential random-graph model framework of Shalizi & Rinaldo (2013). Both follow the graphon construction

$$U_i \overset{\mathrm{iid}}{\sim} \mathrm{Unif}(0,1), \tag{37}$$

$$A_{ij}|U_i, U_j \sim \mathrm{Bernoulli}(W(U_i, U_j)), \tag{38}$$

where $W : [0,1]^2 \to [0,1]$ maps pairs of latent positions to edge probabilities. For the exponential-decay graphon, this mapping depends directly on the distance between latent positions,

$$W_{\exp}(U_i, U_j) = \exp(-\beta|U_i - U_j|), \tag{39}$$

such that connectivity decreases smoothly with latent distance. For the SBM graphon, the latent space is instead partitioned into K blocks through a mapping $b : [0,1] \to \{1,\dots,K\}$. Edge probabilities are then determined by the corresponding entries of an assortative block probability matrix P, yielding the step-function graphon

$$W_{\mathrm{SBM}}(U_i, U_j) = P_{b(U_i),b(U_j)}. \tag{40}$$

Both DGPs define balanced seven-class node-classification problems in which labels and features depend on the latent position, with

$$Y_i \mid U_i \sim P_Y(\cdot \mid U_i), \tag{41}$$

$$X_i \mid U_i \sim P_X(\cdot \mid U_i). \tag{42}$$

For the exponential-decay graphon, class probabilities vary smoothly over the latent space, with greater label ambiguity near transitions between neighboring classes. For the SBM graphon, class probabilities follow the block structure but vary smoothly within blocks, with greater ambiguity near block transitions.

Table 2: Parameters used for the synthetic graph DGPs.

Component	Parameter	Value
Generation	Graph sizes n	$500, 910, 1657, 3017, 5493, 10000$
	Independent sequences	20
Exponential graphon	Decay β	15
SBM graphon	Blocks K	7
	Between-block probability p_{out}	0.02
	Within-block increment δ	0.83
	Block-probability noise σ_P	0.02
Features	Dimension d	12
	Continuous-feature noise σ_X	0.04
	Binary-feature flip probability	0.20
Exponential labels	Gaussian width σ_Y	0.03
SBM labels	Center / boundary class probability	$0.90/0.65$
	Boundary width	0.15
SBM features	Center / boundary class probability	$0.80/0.55$
	Boundary width	0.15

Each node additionally receives 12 noisy continuous and binary features that are informative of the latent class structure without deterministically identifying the observed label. Feature informativeness is likewise allowed to vary with latent position, such that graph structure and node features provide complementary and imperfect information about the class.

Complete generator specifications and parameter settings are provided in Table 2. Tables 3–4 report the realized structural properties of the generated graph sequences across 20 independent realizations.

Each projective graphon sequence comprises six graph sizes, logarithmically spaced between 500 and 10,000 nodes. We first generate the largest graph and construct each smaller graph as the induced subgraph on its first n nodes, thereby yielding a nested sequence from the same underlying realization.

C Local Bootstrap

We construct bootstrap graph replicates using a non-parametric procedure adapted from the local probability estimator of Jeong & Donnat (2025), developed under an exchangeable graphon model and exploiting node similarity. The procedure preserves the observed node set, features and labels across replicates, while resampling the graph structure from smoothed edge probabilities.

For each node $i \in \mathcal{V}$, we first construct a set $\mathcal{N}_k(i)$ of its k most similar nodes according to a chosen node similarity metric. In the synthetic graphon experiments, where the latent coordinates U_i are known, we use Euclidean distance in latent space,

$$d(i,j) = \|U_i - U_j\|_2, \qquad i \neq j. \tag{43}$$

For the OOD benchmark datasets, where latent coordinates are unavailable, we instead use the open-neighborhood Jaccard similarity suggested by Jeong & Donnat (2025),

$$J(i,j) = \frac{|\mathcal{N}(i) \cap \mathcal{N}(j)|}{|\mathcal{N}(i) \cup \mathcal{N}(j)|}, \qquad i \neq j, \tag{44}$$

where $\mathcal{N}(i)$ denotes the open graph neighborhood of node i. This criterion favors nodes with similar observed connectivity profiles, so that the selected neighbors provide local substitutes for node i when estimating its edge probabilities.

The directed edge probability between nodes i and j is then estimated by averaging the observed adjacency of j over $\mathcal{N}_k(i)$:

$$\hat{p}_{ij} = \frac{1}{|\mathcal{N}_k(i)|} \sum_{u \in \mathcal{N}_k(i)} \mathbf{A}_{uj}. \tag{45}$$

Algorithm 1 Conditional local graph bootstrap

Require: Graph $\mathcal{G} = (A, X)$, neighborhood size k, pairwise node similarity metric d, number of replicates B
Ensure: Bootstrap graphs $\{\mathcal{G}_b^{\text{boot}} = (A_b^{\text{boot}}, X)\}_{b=1}^{B}$

for $i \in \mathcal{V}$ **do**
  Construct $\mathcal{N}_k(i)$ as the k nearest nodes similar to i according to d
end for
for $i, j \in \mathcal{V}, i \neq j$ **do**
  $\hat{p}_{ij} \leftarrow \dfrac{1}{|\mathcal{N}_k(i)|} \displaystyle\sum_{u \in \mathcal{N}_k(i)} A_{uj}$
end for
for $i < j$ **do**
  $\hat{p}_{ij}^{\text{sym}} \leftarrow \dfrac{\hat{p}_{ij} + \hat{p}_{ji}}{2}$
end for
for $b = 1, \ldots, B$ **do**
  for $i < j$ **do**
    $A_{ij,b}^{\text{boot}} \sim \text{Bernoulli}\left(\hat{p}_{ij}^{\text{sym}}\right)$
    $A_{ji,b}^{\text{boot}} \leftarrow A_{ij,b}^{\text{boot}}$
  end for
  $A_{ii,b}^{\text{boot}} \leftarrow 0$ for all $i \in \mathcal{V}$
  $\mathcal{G}_b^{\text{boot}} \leftarrow (A_b^{\text{boot}}, X)$
end for
return $\{\mathcal{G}_b^{\text{boot}} = (A_b^{\text{boot}}, X)\}_{b=1}^{B}$.

Thus, $\hat{p}_{ij}$ is the empirical fraction of nodes locally similar to i that are adjacent to j in the observed graph. Because this estimate is generally asymmetric, we obtain an undirected edge probability by

$$\hat{p}_{ij}^{\text{sym}} = \frac{1}{2}(\hat{p}_{ij} + \hat{p}_{ji}), \tag{46}$$

which preserves the asymptotic unbiasedness of the local probability estimator under the conditions of Jeong & Donnat (2025).

Conditional on the resulting probability matrix, bootstrap graphs are generated independently by sampling each unordered node pair according to

$$A_{ij,b}^{\text{boot}} | \hat{p}_{ij}^{\text{sym}} \overset{\text{iid}}{\sim} \text{Bernoulli}(\hat{p}_{ij}^{\text{sym}}), \qquad i < j, \quad b \in \{1, \ldots, B\}, \tag{47}$$

and setting $A_{ii,b}^{\text{boot}} = 0$. The complete graph resampling procedure is given in Algorithm 1. Following the finite-sample specification of Jeong & Donnat (2025), we use $k = 20$ for both the synthetic graphons and the node classification benchmark datasets.

The statistics of 25 bootstrap graphs are compared with those of the original graphs for the synthetic graphons and node classification benchmark datasets in Tables 3– 4 and 5, respectively.

Table 3: Original and bootstrap graph statistics for the exponential-decay graphon at increasing sizes. Values are mean ± standard deviation across 20 graphon realizations, with bootstrap statistics averaged across 25 replicates per realization.

	$n = 500$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	500	500
Edges Edges	$(1.547 \times 10^4) \pm 212.020$	$(1.530 \times 10^4) \pm 220.126$
Average degree	61.882 ± 0.848	61.206 ± 0.881
Degree std.	10.118 ± 1.385	10.327 ± 1.311
Density	$(1.240 \times 10^{-1}) \pm (1.700 \times 10^{-3})$	$(1.227 \times 10^{-1}) \pm (1.765 \times 10^{-3})$
Local clustering	0.404 ± 0.005	0.394 ± 0.004

Components	1	1
Giant component	500	500
Assortativity	0.416 ± 0.096	0.371 ± 0.093
Node homophily	0.631 ± 0.011	0.623 ± 0.012
Edge homophily	0.620 ± 0.013	0.612 ± 0.013
Transitivity	0.396 ± 0.005	0.386 ± 0.005
Triangle density	$(6.169 \times 10^{-3}) \pm (1.826 \times 10^{-4})$	$(5.889 \times 10^{-3}) \pm (1.837 \times 10^{-4})$

	$n = 910$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	910	910
Edges	$(5.131 \times 10^4) \pm 361.446$	$(5.095 \times 10^4) \pm 362.437$
Average degree	112.768 ± 0.794	111.969 ± 0.797
Degree std.	16.285 ± 1.714	16.549 ± 1.712
Density	$(1.241 \times 10^{-1}) \pm (8.739 \times 10^{-4})$	$(1.232 \times 10^{-1}) \pm (8.763 \times 10^{-4})$
Local clustering	0.404 ± 0.003	0.399 ± 0.003
Components	1	1
Giant component	910	910
Assortativity	0.496 ± 0.082	0.464 ± 0.081
Node homophily	0.632 ± 0.005	0.628 ± 0.005
Edge homophily	0.621 ± 0.005	0.617 ± 0.005
Transitivity	0.396 ± 0.003	0.391 ± 0.003
Triangle density	$(6.170 \times 10^{-3}) \pm (9.183 \times 10^{-5})$	$(6.013 \times 10^{-3}) \pm (9.348 \times 10^{-5})$

	$n = 1657$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	1657	1657
Edges Edges	$(1.705 \times 10^5) \pm (1.041 \times 10^3)$	$(1.697 \times 10^5) \pm (1.034 \times 10^3)$
Average degree	205.798 ± 1.256	204.886 ± 1.249
Degree std.	27.837 ± 2.599	28.184 ± 2.567
Density	$(1.243 \times 10^{-1}) \pm (7.585 \times 10^{-4})$	$(1.237 \times 10^{-1}) \pm (7.540 \times 10^{-4})$
Local clustering	0.404 ± 0.002	0.401 ± 0.002
Components	1	1
Giant component	1657	1657
Assortativity	0.572 ± 0.062	0.548 ± 0.062
Node homophily	0.631 ± 0.004	0.629 ± 0.004
Edge homophily	0.620 ± 0.004	0.618 ± 0.004
Transitivity	0.396 ± 0.003	0.393 ± 0.002
Triangle density	$(6.196 \times 10^{-3}) \pm (8.944 \times 10^{-5})$	$(6.109 \times 10^{-3}) \pm (8.044 \times 10^{-5})$

	$n = 3017$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	3017	3017
Edges	$(5.661 \times 10^5) \pm (2.464 \times 10^3)$	$(5.646 \times 10^5) \pm (2.505 \times 10^3)$
Average degree	375.244 ± 1.634	374.306 ± 1.661
Degree std.	48.265 ± 2.831	48.657 ± 2.799
Density	$(1.244 \times 10^{-1}) \pm (5.417 \times 10^{-4})$	$(1.241 \times 10^{-1}) \pm (5.506 \times 10^{-4})$
Local clustering	0.403 ± 0.002	0.402 ± 0.002
Components	1	1
Giant component	3017	3017
Assortativity	0.609 ± 0.043	0.593 ± 0.044

Node homophily	0.631 ± 0.003	0.630 ± 0.003
Edge homophily	0.620 ± 0.003	0.619 ± 0.003
Transitivity	0.395 ± 0.002	0.395 ± 0.001
Triangle density	$(6.206 \times 10^{-3}) \pm (5.918 \times 10^{-5})$	$(6.170 \times 10^{-3}) \pm (6.114 \times 10^{-5})$

	$n = 5493$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	5493	5493
Edges	$(1.878 \times 10^{6}) \pm (6.496 \times 10^{3})$	$(1.875 \times 10^{6}) \pm (6.503 \times 10^{3})$
Average degree	683.611 ± 2.365	682.675 ± 2.368
Degree std.	85.005 ± 4.412	85.439 ± 4.400
Density	$(1.245 \times 10^{-1}) \pm (4.306 \times 10^{-4})$	$(1.243 \times 10^{-1}) \pm (4.311 \times 10^{-4})$
Local clustering	0.404 ± 0.002	0.403 ± 0.001
Components	1	1
Giant component	5493	5493
Assortativity	0.632 ± 0.035	0.622 ± 0.035
Node homophily	0.631 ± 0.003	0.631 ± 0.003
Edge homophily	0.620 ± 0.003	0.619 ± 0.003
Transitivity	0.397 ± 0.002	0.395 ± 0.001
Triangle density	$(6.236 \times 10^{-3}) \pm (5.362 \times 10^{-5})$	$(6.199 \times 10^{-3}) \pm (4.463 \times 10^{-5})$

	$n = 10000$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	10000	10000
Edges	$(6.224 \times 10^{6}) \pm (1.645 \times 10^{4})$	$(6.220 \times 10^{6}) \pm (1.655 \times 10^{4})$
Average degree	1244.879 ± 3.290	1243.939 ± 3.310
Degree std.	152.286 ± 6.828	152.737 ± 6.834
Density	$(1.245 \times 10^{-1}) \pm (3.291 \times 10^{-4})$	$(1.244 \times 10^{-1}) \pm (3.310 \times 10^{-4})$
Local clustering	0.404 ± 0.002	0.403 ± 0.002
Components	1	1
Giant component	10000	10000
Assortativity	0.652 ± 0.031	0.646 ± 0.031
Node homophily	0.631 ± 0.002	0.631 ± 0.002
Edge homophily	0.620 ± 0.002	0.620 ± 0.002
Transitivity	0.396 ± 0.002	0.396 ± 0.001
Triangle density	$(6.222 \times 10^{-3}) \pm (4.641 \times 10^{-5})$	$(6.213 \times 10^{-3}) \pm (3.253 \times 10^{-5})$

Table 4: Original and bootstrap graph statistics for the SBM graphon at increasing sizes. Values are mean $\pm$ standard deviation across 20 graphon realizations, with bootstrap statistics averaged across 25 replicates per realization.

	$n = 500$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	500	500
Edges	$(1.755 \times 10^{4}) \pm 161.030$	$(1.738 \times 10^{4}) \pm 169.203$
Average degree	70.203 ± 0.644	69.502 ± 0.677
Degree std.	8.206 ± 1.813	8.799 ± 1.683
Density	$(1.407 \times 10^{-1}) \pm (1.291 \times 10^{-3})$	$(1.393 \times 10^{-1}) \pm (1.356 \times 10^{-3})$
Local clustering	0.643 ± 0.005	0.550 ± 0.008
Components	1	1

Giant component	500	500
Assortativity	0.570 ± 0.127	0.458 ± 0.125
Node homophily	0.620 ± 0.016	0.601 ± 0.016
Edge homophily	0.619 ± 0.016	0.600 ± 0.017
Transitivity	0.641 ± 0.007	0.549 ± 0.010
Triangle density	$(1.270 \times 10^{-2}) \pm (2.959 \times 10^{-4})$	$(1.070 \times 10^{-2}) \pm (3.086 \times 10^{-4})$

	$n = 910$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	910	910
Edges	$(5.804 \times 10^4) \pm 237.804$	$(5.767 \times 10^4) \pm 240.446$
Average degree	127.554 ± 0.523	126.742 ± 0.528
Degree std.	11.021 ± 2.066	11.708 ± 1.985
Density	$(1.403 \times 10^{-1}) \pm (5.750 \times 10^{-4})$	$(1.394 \times 10^{-1}) \pm (5.814 \times 10^{-4})$
Local clustering	0.644 ± 0.002	0.589 ± 0.006
Components	1	1
Giant component	910	910
Assortativity	0.576 ± 0.116	0.501 ± 0.115
Node homophily	0.612 ± 0.011	0.603 ± 0.011
Edge homophily	0.612 ± 0.010	0.603 ± 0.011
Transitivity	0.641 ± 0.004	0.586 ± 0.007
Triangle density	$(1.262 \times 10^{-2}) \pm (1.232 \times 10^{-4})$	$(1.142 \times 10^{-2}) \pm (1.594 \times 10^{-4})$

	$n = 1657$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	1657	1657
Edges	$(1.926 \times 10^5) \pm 555.421$	$(1.919 \times 10^5) \pm 563.833$
Average degree	232.496 ± 0.670	231.675 ± 0.681
Degree std.	15.500 ± 3.951	16.232 ± 3.744
Density	$(1.404 \times 10^{-1}) \pm (4.048 \times 10^{-4})$	$(1.399 \times 10^{-1}) \pm (4.110 \times 10^{-4})$
Local clustering	0.644 ± 0.002	0.613 ± 0.004
Components	1	1
Giant component	1657	1657
Assortativity	0.575 ± 0.138	0.520 ± 0.137
Node homophily	0.607 ± 0.008	0.603 ± 0.007
Edge homophily	0.607 ± 0.008	0.602 ± 0.007
Transitivity	0.642 ± 0.003	0.612 ± 0.004
Triangle density	$(1.266 \times 10^{-2}) \pm (1.082 \times 10^{-4})$	$(1.200 \times 10^{-2}) \pm (1.083 \times 10^{-4})$

	$n = 3017$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	3017	3017
Edges	$(6.388 \times 10^5) \pm 915.115$	$(6.375 \times 10^5) \pm 954.026$
Average degree	423.445 ± 0.607	422.636 ± 0.632
Degree std.	23.506 ± 4.586	24.242 ± 4.521
Density	$(1.404 \times 10^{-1}) \pm (2.011 \times 10^{-4})$	$(1.401 \times 10^{-1}) \pm (2.097 \times 10^{-4})$
Local clustering	0.644 ± 0.003	0.627 ± 0.003
Components	1	1
Giant component	3017	3017
Assortativity	0.638 ± 0.079	0.594 ± 0.086
Node homophily	0.612 ± 0.006	0.610 ± 0.006

Edge homophily	0.611 ± 0.006	0.609 ± 0.006
Transitivity	0.642 ± 0.003	0.626 ± 0.003
Triangle density	$(1.266 \times 10^{-2}) \pm (7.382 \times 10^{-5})$	$(1.231 \times 10^{-2}) \pm (7.212 \times 10^{-5})$

	$n = 5493$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	5493	5493
Edges	$(2.118 \times 10^{6}) \pm (2.087 \times 10^{3})$	$(2.116 \times 10^{6}) \pm (2.135 \times 10^{3})$
Average degree	771.060 ± 0.760	770.270 ± 0.777
Degree std.	37.061 ± 6.232	37.786 ± 6.088
Density	$(1.404 \times 10^{-1}) \pm (1.384 \times 10^{-4})$	$(1.403 \times 10^{-1}) \pm (1.415 \times 10^{-4})$
Local clustering	0.644 ± 0.002	0.634 ± 0.002
Components	1	1
Giant component	5493	5493
Assortativity	0.690 ± 0.052	0.658 ± 0.059
Node homophily	0.616 ± 0.004	0.615 ± 0.004
Edge homophily	0.616 ± 0.004	0.615 ± 0.004
Transitivity	0.642 ± 0.003	0.633 ± 0.002
Triangle density	$(1.268 \times 10^{-2}) \pm (5.112 \times 10^{-5})$	$(1.247 \times 10^{-2}) \pm (3.925 \times 10^{-5})$

	$n = 10000$	
Metric	Original	Bootstrap
Features	12	12
Classes	7	7
Nodes	10000	10000
Edges	$(7.019 \times 10^{6}) \pm (3.921 \times 10^{3})$	$(7.015 \times 10^{6}) \pm (3.931 \times 10^{3})$
Average degree	1403.808 ± 0.784	1403.021 ± 0.786
Degree std.	59.618 ± 7.229	60.364 ± 7.116
Density	$(1.404 \times 10^{-1}) \pm (7.843 \times 10^{-5})$	$(1.403 \times 10^{-1}) \pm (7.863 \times 10^{-5})$
Local clustering	0.644 ± 0.003	0.638 ± 0.001
Components	1	1
Giant component	10000	10000
Assortativity	0.725 ± 0.027	0.703 ± 0.029
Node homophily	0.618 ± 0.004	0.617 ± 0.004
Edge homophily	0.617 ± 0.004	0.617 ± 0.004
Transitivity	0.641 ± 0.002	0.637 ± 0.001
Triangle density	$(1.266 \times 10^{-2}) \pm (3.880 \times 10^{-5})$	$(1.257 \times 10^{-2}) \pm (2.195 \times 10^{-5})$

Table 5: Original and bootstrap graph statistics for OOD datasets. Values are bootstrap means $\pm$ standard deviations across 25 bootstrap replicates.

	CoraML		PubMed	
Metric	Original	Bootstrap	Original	Bootstrap
Features	2879	2879	500	500
Classes	7	7	3	3
Nodes	2995	2995	19717	19717
Edges	8158	9802.200 ± 82.619	44324	60935.400 ± 243.491
Average degree	5.450	6.546 ± 0.055	4.500	6.181 ± 0.025
Degree std.	8.240	7.458 ± 0.132	7.430	8.509 ± 0.049
Density	1.820×10^{-3}	$(2.186 \times 10^{-3}) \pm (1.843 \times 10^{-5})$	2.280×10^{-4}	$(3.135 \times 10^{-4}) \pm (1.253 \times 10^{-6})$
Local clustering	0.285	0.062 ± 0.004	0.061	0.027 ± 0.003
Components	61	66.360 ± 6.557	1	1278.800 ± 37.791
Giant component	2810	2928.680 ± 6.638	19717	18415.600 ± 38.176
Assortativity	-0.072	-0.062 ± 0.004	-0.044	-0.038 ± 0.005

Node homophily	0.810	0.560 ± 0.005	0.792	0.626 ± 0.002
Edge homophily	0.789	0.595 ± 0.005	0.802	0.696 ± 0.002
Transitivity	0.116	0.046 ± 0.002	0.054	0.025 ± 0.001
Triangle density	1.190×10^{-6}	$(4.712 \times 10^{-7}) \pm (1.986 \times 10^{-8})$	9.840×10^{-9}	$(6.590 \times 10^{-9}) \pm (2.318 \times 10^{-10})$

	CiteSeer		AmazonComputers	
Metric	Original	Bootstrap	Original	Bootstrap
Features	3703	3703	767	767
Classes	6	6	10	10
Nodes	3327	3327	13752	13752
Edges	4552	5887.080 ± 97.053	245861	268828.640 ± 434.954
Average degree	2.740	3.539 ± 0.058	35.800	39.097 ± 0.063
Degree std.	3.380	4.133 ± 0.092	70.300	66.104 ± 0.153
Density	8.230×10^{-4}	$(1.064 \times 10^{-3}) \pm (1.754 \times 10^{-5})$	2.600×10^{-3}	$(2.843 \times 10^{-3}) \pm (4.600 \times 10^{-6})$
Local clustering	0.150	0.019 ± 0.002	0.362	0.190 ± 0.003
Components	438	362.360 ± 24.088	314	9.640 ± 3.187
Giant component	2120	2931.480 ± 28.849	13381	13743.360 ± 3.187
Assortativity	0.048	0.041 ± 0.013	-0.057	-0.052 ± 0.001
Node homophily	0.706	0.333 ± 0.005	0.785	0.733 ± 0.001
Edge homophily	0.736	0.447 ± 0.005	0.777	0.746 ± 0.001
Transitivity	0.131	0.036 ± 0.002	0.108	0.090 ± 0.001
Triangle density	1.910×10^{-7}	$(8.483 \times 10^{-8}) \pm (7.211 \times 10^{-9})$	3.550×10^{-6}	$(2.790 \times 10^{-6}) \pm (3.004 \times 10^{-8})$

	AmazonPhotos		CoauthorCS	
Metric	Original	Bootstrap	Original	Bootstrap
Features	745	745	6805	6805
Classes	8	8	15	15
Nodes	7650	7650	18333	18333
Edges	119081	133971.160 ± 265.568	81894	91787.760 ± 266.237
Average degree	31.100	35.025 ± 0.069	8.930	10.013 ± 0.029
Degree std.	47.300	45.657 ± 0.125	9.110	8.149 ± 0.043
Density	4.070×10^{-3}	$(4.579 \times 10^{-3}) \pm (9.077 \times 10^{-6})$	4.870×10^{-4}	$(5.462 \times 10^{-4}) \pm (1.584 \times 10^{-6})$
Local clustering	0.406	0.254 ± 0.003	0.356	0.056 ± 0.002
Components	136	6.400 ± 2.273	1	55.480 ± 7.434
Giant component	7487	7644.600 ± 2.273	18333	18278.240 ± 7.683
Assortativity	-0.045	-0.031 ± 0.001	0.113	0.088 ± 0.004
Node homophily	0.836	0.782 ± 0.001	0.832	0.638 ± 0.002
Edge homophily	0.827	0.794 ± 0.001	0.808	0.665 ± 0.001
Transitivity	0.178	0.156 ± 0.001	0.185	0.062 ± 0.001
Triangle density	9.650×10^{-6}	$(8.917 \times 10^{-6}) \pm (8.917 \times 10^{-8})$	8.470×10^{-8}	$(2.901 \times 10^{-8}) \pm (5.275 \times 10^{-10})$

	CoauthorPhysics		OGBN-ArXiv	
Metric	Original	Bootstrap	Original	Bootstrap
Features	8415	8415	128	128
Classes	5	5	40	40
Nodes	34493	34493	169343	169343
Edges	247962	266575.000 ± 462.087	1.158×10^6	$(1.902 \times 10^6) \pm (1.061 \times 10^3)$
Average degree	14.400	15.457 ± 0.027	13.700	22.469 ± 0.013
Degree std.	15.600	13.170 ± 0.042	68.600	78.516 ± 0.117
Density	4.170×10^{-4}	$(4.481 \times 10^{-4}) \pm (7.768 \times 10^{-7})$	8.070×10^{-5}	$(1.327 \times 10^{-4}) \pm (7.403 \times 10^{-8})$
Local clustering	0.373	0.076 ± 0.001	0.235	0.060 ± 0.002
Components	1	28.000 ± 4.933	1	1123.560 ± 28.833
Giant component	34493	34465.960 ± 4.945	169343	168214.880 ± 29.359
Assortativity	0.201	0.213 ± 0.003	-0.043	-0.056 ± 0.000
Node homophily	0.915	0.811 ± 0.001	0.635	0.429 ± 0.000
Edge homophily	0.931	0.862 ± 0.001	0.654	0.435 ± 0.000
Transitivity	0.189	0.073 ± 0.001	0.016	0.008 ± 0.000

Triangle density	6.900×10^{-8}	$(2.421 \times 10^{-8}) \pm (3.094 \times 10^{-10})$	2.740×10^{-9}	$(1.893 \times 10^{-9}) \pm (5.433 \times 10^{-11})$

D Model Architectures and Parameters

D.1 Graph EDL Models

For the graph EDL baselines, we consider S-BGCN-K and S-BGCN-T-K (Zhao et al., 2020), GPN (Stadler et al., 2021), GPN-GD$_{\text{EVI}}$ and GPN-GD$_{\text{LAT}}$ (Hart et al., 2023), GPN-LOP (Damke & Hüllermeier, 2024a), and CUQ-GCN, CUQ-GAT, and CUQ-APPNP (Damke & Hüllermeier, 2024b). Unless otherwise specified, we use the same hyperparameters across the graphon information-growth experiments, while for the OOD benchmarks we use the dataset-specific configurations suggested in the original implementations.

GPN, GPN-GD, GPN-LOP, CUQ-GNN. GPN, all GPN-GD variants, GPN-LOP, and all CUQ-GNN variants use a learning rate of 0.01 across all datasets and 10 radial-flow layers for density estimation. For GPN, all GPN-GD variants, GPN-LOP, and CUQ-APPNP, APPNP propagation uses $K = 10$. CUQ-GCN, CUQ-GAT, and CUQ-APPNP differ only in the graph convolution used before density estimation; CUQ-GAT uses one attention head per layer, negative slope 0.2, and no attention-coefficient dropout. GPN-GD uses the same hyperparameters as GPN, and additionally tunes the graph distance regularization weight and activation function, using the same hyperparameters for both variants. Graphon and OOD dataset-specific hyperparameters are reported in Tables 6 and 7.

S-BGCN. We evaluate S-BGCN-K and S-BGCN-T-K. We use a learning rate of 0.01 across datasets. S-BGCN-K uses λ_1 to weight the KL regularization term, while S-BGCN-T-K additionally uses λ_2 to weight the teacher regularization term. The graphon and OOD dataset-specific hyperparameters are reported in Table 8. Uncertainty estimates are computed using 100 MC dropout samples.

D.2 G-Ens and GB-Ens

Unless otherwise specified, G-Ens and GB-Ens use the same hyperparameters for the graphons, while for the OOD benchmarks we perform the same dataset-specific hyperparameter search, selecting model and training hyperparameters by validation accuracy; GB-Ens additionally applies graph bootstrapping.

APPNP backbone. For the graphon experiments, we use an APPNP architecture (Gasteiger et al., 2019) consisting of a two-layer MLP with hidden dimension 64, dropout 0.3, followed by APPNP propagation with $K = 10$ iterations and teleport probability $\alpha = 0.1$. We use a learning rate of 0.01 and weight decay 5×10^{-4}. For the OOD datasets, we fix the MLP to two layers and $K = 10$, and search over hidden dimensions in $\{64, 128, 256\}$; dropout rates in $\{0.3, 0.5, 0.6\}$; teleport probabilities in $\{0.05, 0.1, 0.2\}$; learning rates in $\{0.005, 0.01, 0.02\}$; and weight decay in $\{10^{-5}, 5 \times 10^{-4}, 5 \times 10^{-3}\}$.

GAT backbone. For the graphon experiments, we use a two-layer GAT (Veličković et al., 2018) with total hidden dimension 64, 8 attention heads, feature and attention-coefficient dropout rates of 0.3 and 0.2, respectively, attention negative slope of 0.2, learning rate of 0.005, and weight decay of 5×10^{-4}. For the OOD datasets, we search over 2 or 3 GAT layers; total hidden dimensions in $\{64, 128, 256\}$; 4 or 8 attention heads; feature dropout rates in $\{0.2, 0.3, 0.5, 0.6\}$; attention-coefficient dropout rates in $\{0.3, 0.5, 0.6\}$; learning rates in $\{0.005, 0.01\}$; and weight decay in $\{10^{-4}, 10^{-3}\}$. The attention negative slope is fixed to 0.2.

GCN backbone. For the graphon experiments, we use a two-layer GCN (Kipf & Welling, 2017) with hidden dimension 64, dropout of 0.3, learning rate of 0.01, and weight decay of 5×10^{-4}. For the OOD datasets, we search over 2 or 3 GCN layers; hidden dimensions in $\{64, 128, 256\}$; dropout rates in $\{0.2, 0.4, 0.5\}$; learning rates in $\{0.01, 0.02, 0.03\}$; and weight decay in $\{5 \times 10^{-5}, 5 \times 10^{-4}, 5 \times 10^{-3}\}$.

E Additional Experimental Results

RQ2: Graph EDL models are inconsistent epistemic predictors.

Table 6: Hyperparameters for GPN, GPN-LOP, and CUQ-GNN.

Dataset	Hidden	Latent	Layers	Dropout	APPNP α	Entropy reg.	Weight decay
Graphons	64	16	2	0.50	0.10	10^{-3}	10^{-3}
CoraML	64	16	2	0.50	0.10	10^{-3}	10^{-3}
CiteSeer	64	16	2	0.50	0.10	10^{-3}	10^{-3}
PubMed	64	16	2	0.50	0.10	10^{-3}	10^{-3}
AmazonComputers	64	10	2	0.50	0.20	10^{-5}	5×10^{-4}
AmazonPhotos	64	10	2	0.50	0.20	10^{-5}	5×10^{-4}
CoauthorCS	64	16	2	0.50	0.10	10^{-3}	10^{-3}
CoauthorPhysics	64	16	2	0.50	0.10	10^{-3}	10^{-3}
OGBN-ArXiv	256	16	3	0.25	0.20	10^{-3}	0

Table 7: Hyperparameters for GPN-GD.

Dataset	Entropy reg.	Distance reg.	Activation
Graphons	0	10^{-4}	GELU
CoraML	0	10^{-4}	GELU
CiteSeer	10^{-4}	$10^{-9.5}$	LogSigmoid
PubMed	10^{-5}	10^{-4}	ReLU
AmazonComputers	10^{-5}	10^{-4}	ReLU
AmazonPhotos	10^{-5}	10^{-11}	ReLU
CoauthorCS	0	10^{-6}	ReLU
CoauthorPhysics	10^{-4}	$10^{-4.5}$	LogSigmoid
OGBN-ArXiv	10^{-5}	10^{-8}	ReLU

The SBM results in Figures 4a and 4b reproduce the patterns observed for the exponential-decay graphon: EDL methods fail to exhibit consistent EU concentration, while AU remains sensitive to the regularization parameter λ. Similar differences in EU magnitude across methods persist on real-world datasets, as shown in Table 9. For completeness, Table 10 reports predictive accuracy averaged over the two graphon sequences; GPN variants share a common parameterization, as in their original formulations, resulting in lower accuracy for some variants, particularly GPN-GD$_{\text{EVI}}$.

RQ3: GB-Ens is a family of candidate consistent epistemic predictors.

Complementing the results reported in the main text, Figure 5a shows that on the SBM, GB-Ens EU decreases from $N = 500$ to $N = 10,000$ by 92% for APPNP and 94% for GCN, while GAT exhibits a weaker, non-monotonic EU reduction of 34%, reaching a maximum of 42% at an intermediate graph size. Figure 5b further decomposes EU into seed and bootstrap contributions of similar magnitude, both decreasing with increasing information. For GAT, the non-monotonicity is driven primarily by the seed component, whereas the bootstrap component follows a more consistently decreasing trajectory. Regarding aleatoric uncertainty, Figure 6b shows that GB-Ens AU decreases and eventually stabilizes, with all three backbones approaching similar plateaus.

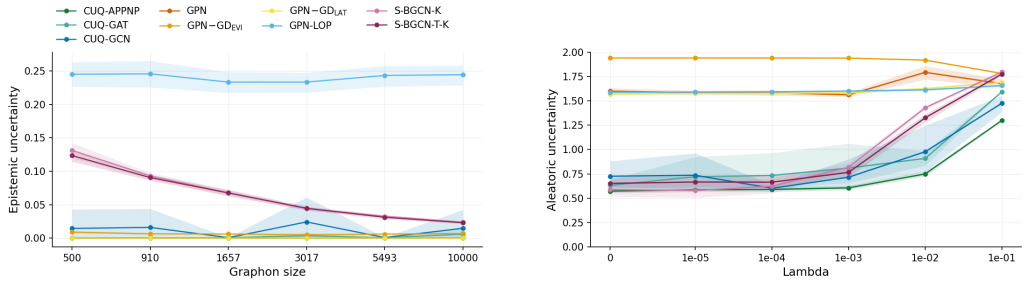
Figure 6a shows that G-Ens similarly exhibit no consistent EU contraction on the SBM, while Figure 6b shows strongly backbone-dependent AU trajectories. For completeness, Table 12 reports predictive accuracy averaged over the two graphon information-growth sequences for G-Ens and GB-Ens. For the true-DGP resampling diagnostic, we compute normalized node-wise Wasserstein-1 distances on a random 10% subset of test nodes and average them across nodes and 20 projective-sequence replicates, using the Euclidean diameter $\sqrt{2}$ of the probability simplex for normalization. Table 11 shows that, on the SBM, distances remain relatively low for GCN (0.12–0.14) and APPNP (0.20–0.34), whereas GAT exhibits larger distances ranging from 0.24 to 0.45.

F Out-Of-Distribution Detection

We report the full OOD detection results for the EDL baselines and, for completeness, extend the evaluation to G-Ens and GB-Ens in Tables 13 to 20, for which we similarly find that the most effective uncertainty type depends on the model and OOD protocol. Bold highlights the best-performing uncertainty type(s) for each model and shift. GB-Ens bootstrap and original dataset statistics are reported in Table 5. The observed discrepancies in graph statistics indicate that the local bootstrap

Table 8: Hyperparameters for S-BGCN-K and S-BGCN-T-K.

Dataset	Hidden	Dropout	L_2 reg.	λ_1	λ_2
Graphons	16	0.50	5×10^{-4}	10^{-3}	1
CoraML	16	0.50	5×10^{-4}	10^{-3}	1
CiteSeer	16	0.50	5×10^{-4}	10^{-3}	1
PubMed	16	0.50	5×10^{-4}	10^{-3}	1
AmazonComputers	64	0.20	10^{-4}	10^{-3}	1
AmazonPhotos	64	0.20	10^{-4}	10^{-3}	1
CoauthorCS	64	0.10	10^{-3}	10^{-3}	1
CoauthorPhysics	64	0.10	10^{-3}	10^{-3}	1
OGBN-ArXiv	64	0.50	10^{-4}	10^{-4}	1



(a) Epistemic uncertainty under information growth.

(b) Aleatoric uncertainty across λ at $N = 10{,}000$.

Figure 4: Uncertainty estimates of EDL models on the SBM.

does not preserve all topological properties of the observed graphs, limiting its fidelity as an approximation of graph observation variability; nevertheless, in these downstream OOD experiments uncertainty serves only as a detection score, and this structural degradation does not automatically translate into degraded OOD detection performance. We follow the leave-out-classes (LOC) and feature isolation (FI) protocols of Stadler et al. (2021); Damke & Hüllermeier (2024a;b).

On OGBN-ArXiv, CUQ-GAT and CUQ-GCN achieve very low accuracy (≈ 0.051–0.073) using the hyperparameters recommended in the original paper (Damke & Hüllermeier, 2024b), while GPN-LOP exceeds the available 80 GB of NVIDIA A100 GPU memory; we therefore do not report results for these models.

Datasets. We use CoraML (McCallum et al., 2000), CiteSeer (Giles et al., 1998), PubMed (Namata et al., 2012), CoauthorCS, CoauthorPhysics, AmazonComputers, AmazonPhotos (Shchur et al., 2018), and OGBN-ArXiv (Hu et al., 2020).

Leave-out-classes (LOC). Following Stadler et al. (2021), we partition the label space $\mathcal{Y}$ into retained classes $\mathcal{Y}_{\text{ID}}$ and the same left-out classes $\mathcal{Y}_{\text{OOD}}$ used in their LOC setup. Nodes from $\mathcal{Y}_{\text{OOD}}$ are excluded from training and validation but remain in $\mathcal{G}$, preserving the transductive setting. For GB-Ens, the same class partition is fixed across bootstrap replicates, with each ensemble member evaluated on its corresponding bootstrap graph.

Feature isolation (FI). We sample $n_{\text{iso}} = 100$ test nodes and, for each node i, construct a copy of $\mathcal{G}$ in which only X_i is replaced by coordinate-wise Gaussian $\mathcal{N}(0, \sigma^2 I_D)$ or Bernoulli $\text{Bernoulli}(0.5)$ noise. For GB-Ens, the same nodes and feature perturbations are used across bootstrap replicates. Each perturbed node is evaluated in its corresponding graph as OOD; for GB-Ens, this evaluation is performed separately on each member's bootstrap graph, while the unperturbed test set $\mathcal{T}^{(\text{te})}$ forms the ID set.

Metrics. Using total, aleatoric, and epistemic uncertainty as OOD scores, we report AUROC, together with ID accuracy and FI OOD accuracy.

Table 9: Mean epistemic uncertainty estimated by EDL methods on OOD datasets.

Method	CoraML	CiteSeer	PubMed	AmazonPhotos	AmazonComputers
CUQ-APPNP	4.42×10^{-4}	6.64×10^{-4}	4.80×10^{-5}	8.00×10^{-6}	4.00×10^{-6}
CUQ-GAT	1.87×10^{-4}	4.07×10^{-4}	2.60×10^{-5}	4.20×10^{-5}	1.66×10^{-4}
CUQ-GCN	7.97×10^{-4}	7.90×10^{-4}	1.79×10^{-4}	2.20×10^{-4}	5.65×10^{-4}
GPN	3.65×10^{-4}	6.21×10^{-4}	5.00×10^{-6}	1.00×10^{-5}	7.00×10^{-6}
GPN-GD$_{\text{EVI}}$	8.89×10^{-2}	5.01×10^{-3}	4.44×10^{-2}	1.76×10^{-5}	1.38×10^{-1}
GPN-GD$_{\text{LAT}}$	2.21×10^{-4}	4.77×10^{-3}	1.93×10^{-6}	8.05×10^{-6}	3.32×10^{-6}
GPN-LOP	3.15×10^{-1}	2.09×10^{-1}	1.81×10^{-1}	2.46×10^{-1}	3.23×10^{-1}
S-BGCN-K	2.69×10^{-2}	7.88×10^{-2}	4.47×10^{-2}	2.66×10^{-3}	1.71×10^{-3}
S-BGCN-T-K	4.30×10^{-2}	6.32×10^{-2}	4.38×10^{-2}	4.36×10^{-3}	2.85×10^{-3}

Method	CoauthorCS	CoauthorPhysics	OGBN-ArXiv
CUQ-APPNP	1.09×10^{-4}	1.44×10^{-4}	1.20×10^{-2}
CUQ-GAT	5.79×10^{-2}	2.10×10^{-5}	4.10×10^{-1}
CUQ-GCN	3.48×10^{-2}	2.08×10^{-4}	2.40×10^{-1}
GPN	4.00×10^{-6}	1.10×10^{-5}	5.44×10^{-3}
GPN-GD$_{\text{EVI}}$	1.02×10^{-2}	5.31×10^{-2}	2.05×10^{-3}
GPN-GD$_{\text{LAT}}$	1.02×10^{-6}	2.72×10^{-5}	1.05×10^{-5}
GPN-LOP	5.33×10^{-1}	2.65×10^{-1}	OOM
S-BGCN-K	1.62×10^{-3}	4.35×10^{-3}	5.33×10^{-2}
S-BGCN-T-K	5.62×10^{-3}	7.06×10^{-3}	5.68×10^{-2}

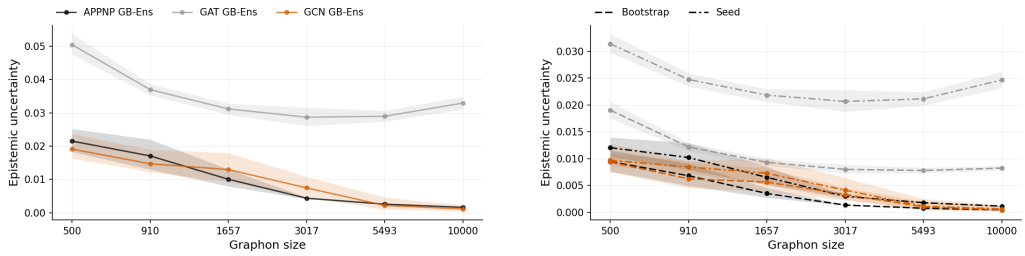
Table 10: EDL test accuracy averaged across each graphon sequence.

DGP	CUQ-APPNP	CUQ-GAT	CUQ-GCN	GPN	GPN-GD$_{\text{EVI}}$	GPN-GD$_{\text{LAT}}$	GPN-LOP	S-BGCN-K	S-BGCN-T-K
Exp.-decay	0.940	0.910	0.848	0.555	0.158	0.761	0.793	0.904	0.906
SBM	0.822	0.832	0.800	0.786	0.159	0.821	0.834	0.823	0.823

G Computational Complexity

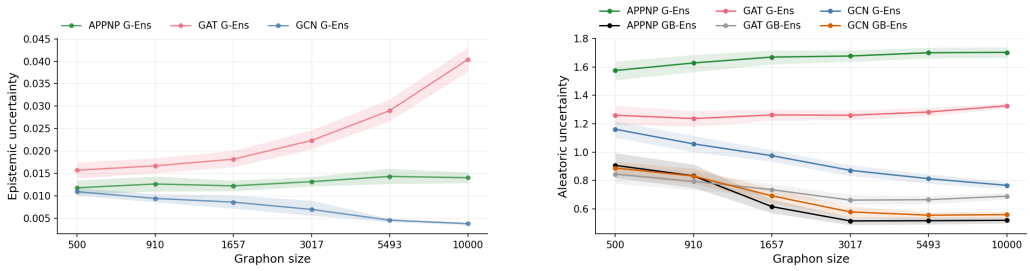
Graph EDL methods can themselves incur substantial computational costs, with different methods trading off preprocessing, training, inference, and memory requirements. Ensemble-based methods require repeated model fitting, but retain manageable per-model computation and peak memory and are naturally parallelizable. GPN and CUQ-GCN admit $O(N+M)$ implementations, whereas GPN-LOP and S-BGCN-K can incur $O(N^2)$ computation due to dense propagation and pairwise GKDE prior construction, respectively. For GCN-based ensembles, training costs $O(N + M)$ per member; GB-Ens additionally requires bootstrap generation, whose cost depends on the k-NN construction and can reach $O(N^2)$ for general similarity search, although cheaper alternatives are available, such as the open Jaccard similarity metric used in our experiments. Bootstrap generation is repeated for each bootstrap sample. Across ensemble members, training cost scales linearly for both GB-Ens and G-Ens but is naturally parallelizable. Moreover, processing members independently and accumulating uncertainty statistics online keeps peak memory at $O(N+M)$, comparable to GPN, CUQ-GNN, and S-BGCN-K, whereas GPN-LOP can require $O(N^2)$ memory due to dense propagation.

Tables 21, 22 and 23 complement the asymptotic analysis with empirical measurements on CoraML, CoauthorPhysics and OGBN-ArXiv, obtained on a dedicated NVIDIA A100 64 GB GPU. GB-Ens uses 100 models in the 25×4 bootstrap–seed decomposition, while G-Ens uses 10 randomly initialized models. Preprocessing time reports the total preprocessing cost required by each method, including generation of all bootstrap samples for GB-Ens. Epoch time is reported per model and training epoch, while inference time reports the total cost of producing one uncertainty estimate, aggregating the 100 forward passes required by GB-Ens and S-BGCN variants and the 10 forward passes required by G-Ens. Peak inference memory is measured when processing ensemble members sequentially. GB-Ens therefore trades increased total preprocessing, training, and inference costs for moderate peak memory requirements: graph bootstrapping becomes particularly costly on larger graphs, while its 100 members can be trained independently and in parallel, and sequential inference limits peak memory. This trade-off differs across EDL baselines: some avoid repeated model fitting and inference, whereas preprocessing-heavy methods such as S-BGCN-K incur comparable preprocessing costs on OGBN-ArXiv, and dense propagation makes GPN-LOP considerably more expensive on CoauthorPhysics and infeasible on OGBN-ArXiv under the available GPU memory.



(a) Total epistemic uncertainty. (b) Bootstrap and training-seed decomposition.

Figure 5: GB-Ens mean epistemic uncertainty on the SBM.



(a) Epistemic uncertainty of G-Ens. (b) Aleatoric uncertainty of GB-Ens and G-Ens.

Figure 6: Ensemble mean uncertainties on the SBM.

Table 11: Mean normalized Wasserstein distances between GB-Ens- and true-DGP-induced second-order distributions.

		Graph size					
Graphon	Backbone	500	910	1657	3017	5493	10000
Exp.-decay	APPNP	0.2991	0.2931	0.2870	0.2886	0.2841	0.2826
	GAT	0.4482	0.5102	0.5505	0.5790	0.5946	0.6156
	GCN	0.2053	0.2113	0.2119	0.2071	0.2059	0.2050
SBM	APPNP	0.2049	0.2185	0.3146	0.3483	0.3471	0.3431
	GAT	0.2440	0.2998	0.3684	0.4130	0.4365	0.4506
	GCN	0.1277	0.1194	0.1437	0.1527	0.1524	0.1480

Table 12: G-Ens and GB-Ens test accuracy averaged across each graphon sequence.

DGP	APPNP G-Ens	GAT G-Ens	GCN G-Ens	APPNP GB-Ens	GAT GB-Ens	GCN GB-Ens
Exp.-decay	0.808	0.731	0.902	0.946	0.931	0.932
SBM	0.704	0.772	0.814	0.816	0.830	0.830

Table 13: OOD detection on CoraML measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.790	**0.761**	**0.761**	0.665	0.080	**1.000**	0.992	**1.000**	0.760	**0.603**	**0.603**	0.596
CUQ-GAT	0.764	**0.697**	**0.697**	0.562	0.070	0.996	0.996	**1.000**	0.740	**0.544**	**0.544**	0.537
CUQ-GCN	0.712	**0.715**	**0.715**	0.616	0.110	**1.000**	0.998	**1.000**	0.700	**0.574**	**0.574**	0.531
GPN	0.802	0.836	0.836	**0.837**	0.790	0.559	0.559	**0.583**	0.800	**0.578**	**0.578**	0.577
GPN-GD$_{\text{EVI}}$	0.782	**0.892**	0.862	0.879	0.760	**0.554**	0.551	0.545	0.760	**0.556**	0.554	0.544
GPN-GD$_{\text{LAT}}$	0.790	**0.832**	**0.832**	0.824	0.790	0.555	0.555	**0.595**	0.790	0.579	0.579	**0.585**
LOP-GPN	0.795	0.837	**0.865**	0.507	0.840	0.726	0.641	**0.836**	0.810	0.668	0.644	**0.673**
S-BGCN-K	0.513	**0.865**	0.864	0.462	0.260	0.028	0.002	**1.000**	0.500	0.571	**0.579**	0.423
S-BGCN-T-K	0.786	0.885	**0.890**	0.368	0.240	0.169	0.009	**1.000**	0.780	0.633	**0.653**	0.362
G-Ens$_{\text{APPNP}}$	0.814	**0.835**	0.831	0.789	0.240	0.272	0.072	**0.668**	0.725	0.667	**0.671**	0.532
G-Ens$_{\text{GAT}}$	0.814	0.850	**0.857**	0.621	0.200	0.217	0.002	**0.863**	0.735	0.633	**0.644**	0.469
G-Ens$_{\text{GCN}}$	0.816	**0.848**	0.847	0.810	0.200	0.226	0.080	**0.540**	0.810	0.670	**0.671**	0.614
GB-Ens$_{\text{APPNP}}$	0.794	**0.807**	0.799	0.726	0.240	0.304	0.025	**0.825**	0.650	0.728	**0.734**	0.499
GB-Ens$_{\text{GAT}}$	0.770	0.797	**0.812**	0.625	0.210	0.389	0.002	**0.989**	0.680	0.658	**0.679**	0.520
GB-Ens$_{\text{GCN}}$	0.687	**0.766**	0.758	0.683	0.280	0.533	0.070	**0.983**	0.730	0.603	**0.626**	0.479

Table 14: OOD detection on CiteSeer measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.647	**0.751**	**0.751**	0.616	0.080	**1.000**	0.711	**1.000**	0.570	**0.640**	**0.640**	0.591
CUQ-GAT	0.613	**0.768**	**0.768**	0.679	0.100	0.941	0.701	**0.989**	0.560	**0.597**	**0.597**	**0.597**
CUQ-GCN	0.557	0.593	**0.594**	0.571	0.110	0.999	0.535	**1.000**	0.480	**0.576**	**0.576**	0.479
GPN	0.678	**0.741**	**0.741**	0.615	0.690	0.489	0.487	**0.702**	0.630	**0.599**	**0.599**	0.569
GPN-GD$_{\text{EVI}}$	0.688	**0.807**	**0.807**	0.774	0.660	0.486	0.482	**0.599**	0.670	**0.568**	**0.568**	0.558
GPN-GD$_{\text{LAT}}$	0.687	**0.805**	**0.805**	0.773	0.670	0.486	0.482	**0.598**	0.680	**0.568**	**0.568**	0.557
LOP-GPN	0.680	**0.748**	0.724	0.596	0.640	0.746	0.596	**0.874**	0.620	0.700	0.620	**0.713**
S-BGCN-K	0.699	**0.739**	0.737	0.413	0.280	0.159	0.002	**1.000**	0.720	0.683	**0.708**	0.415
S-BGCN-T-K	0.723	0.861	**0.876**	0.255	0.260	0.230	0.005	**1.000**	0.740	0.704	**0.731**	0.393
G-Ens$_{\text{APPNP}}$	0.716	**0.805**	0.804	0.545	0.185	0.099	0.002	**0.767**	0.540	0.832	**0.839**	0.238
G-Ens$_{\text{GAT}}$	0.715	**0.799**	0.793	0.565	0.180	0.121	0.005	**0.862**	0.535	0.684	**0.694**	0.458
G-Ens$_{\text{GCN}}$	0.706	**0.829**	**0.829**	0.674	0.210	0.117	0.016	**0.581**	0.720	0.701	**0.702**	0.461
GB-Ens$_{\text{APPNP}}$	0.676	**0.771**	**0.771**	0.399	0.200	0.037	0.000	**0.777**	0.370	0.892	**0.902**	0.125
GB-Ens$_{\text{GAT}}$	0.666	**0.802**	0.788	0.548	0.210	0.075	0.000	**1.000**	0.360	0.764	**0.781**	0.418
GB-Ens$_{\text{GCN}}$	0.667	0.789	**0.794**	0.355	0.160	0.036	0.000	**0.970**	0.410	0.771	**0.812**	0.154

Table 15: OOD detection on PubMed measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.852	**0.693**	**0.693**	0.643	0.300	0.975	0.942	**0.994**	0.770	**0.652**	**0.652**	0.616
CUQ-GAT	0.835	**0.709**	**0.709**	0.618	0.670	0.660	0.659	**0.684**	0.780	**0.586**	**0.586**	0.540
CUQ-GCN	0.845	**0.717**	**0.717**	0.591	0.340	0.936	0.858	**0.945**	0.770	**0.613**	**0.613**	0.554
GPN	0.837	0.730	0.730	**0.769**	0.820	0.560	0.560	**0.642**	0.820	0.593	0.593	**0.638**
GPN-GD$_{\text{EVI}}$	0.827	0.800	0.771	**0.857**	0.820	0.578	0.569	**0.621**	0.820	0.585	0.579	**0.615**
GPN-GD$_{\text{LAT}}$	0.842	0.724	0.724	**0.764**	0.810	0.559	0.559	**0.658**	0.800	0.596	0.596	**0.640**
LOP-GPN	0.834	0.748	**0.781**	0.651	0.780	0.696	0.588	**0.767**	0.770	0.578	0.446	**0.689**
S-BGCN-K	0.848	0.625	**0.627**	0.500	0.400	0.342	0.001	**1.000**	0.770	0.646	**0.679**	0.288
S-BGCN-T-K	0.850	0.607	**0.611**	0.507	0.400	0.447	0.006	**1.000**	0.790	0.654	**0.677**	0.308
G-Ens$_{\text{APPNP}}$	0.867	0.663	**0.664**	0.641	0.385	0.138	0.028	**0.463**	0.730	**0.781**	0.780	0.540
G-Ens$_{\text{GAT}}$	0.845	**0.698**	0.697	0.657	0.365	0.397	0.034	**0.679**	0.765	0.648	**0.652**	0.560
G-Ens$_{\text{GCN}}$	0.854	0.728	0.728	**0.733**	0.400	0.192	0.083	**0.375**	0.820	**0.696**	**0.696**	0.604
GB-Ens$_{\text{APPNP}}$	0.857	0.647	**0.648**	0.600	0.380	0.175	0.025	**0.536**	0.640	**0.775**	0.773	0.551
GB-Ens$_{\text{GAT}}$	0.845	**0.676**	0.675	0.657	0.320	0.573	0.029	**0.962**	0.730	0.661	**0.668**	0.576
GB-Ens$_{\text{GCN}}$	0.829	0.618	**0.626**	0.568	0.400	0.261	0.060	**0.780**	0.720	0.641	**0.649**	0.514

Table 16: OOD detection on Amazon Computers measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.838	**0.717**	**0.717**	0.589	0.010	**1.000**	**1.000**	**1.000**	0.720	**0.698**	**0.698**	0.543
CUQ-GAT	0.875	**0.773**	**0.773**	0.719	0.170	0.908	0.907	**0.909**	0.820	**0.503**	**0.503**	0.484
CUQ-GCN	0.831	**0.632**	**0.632**	0.538	0.210	0.896	0.895	**0.942**	0.790	**0.526**	**0.526**	0.494
GPN	0.810	0.749	0.749	**0.762**	0.720	0.551	0.551	**0.562**	0.730	0.554	0.554	**0.560**
GPN-GD$_{EVI}$	0.381	0.408	0.382	**0.511**	0.910	0.515	0.515	**0.582**	0.430	0.497	**0.505**	0.482
GPN-GD$_{LAT}$	0.832	**0.768**	**0.768**	0.607	0.900	0.515	0.515	**0.580**	0.800	0.567	0.567	**0.583**
LOP-GPN	0.792	0.825	**0.876**	0.507	0.720	**0.891**	0.818	0.871	0.740	0.713	0.637	**0.717**
S-BGCN-K	0.396	0.752	**0.753**	0.593	0.400	0.235	0.164	**0.960**	0.440	0.491	0.491	**0.509**
S-BGCN-T-K	0.626	0.785	**0.787**	0.581	0.320	0.140	0.078	**0.980**	0.660	**0.507**	0.506	0.502
G-Ens$_{APPNP}$	0.875	0.773	**0.774**	0.751	0.140	0.359	0.064	**0.580**	0.785	**0.737**	**0.737**	0.697
G-Ens$_{GAT}$	0.852	0.810	**0.817**	0.760	0.170	0.866	0.120	**0.980**	0.830	0.537	**0.538**	0.511
G-Ens$_{GCN}$	0.854	0.727	0.722	**0.781**	0.370	0.501	0.375	**0.759**	0.820	0.507	**0.508**	0.476
GB-Ens$_{APPNP}$	0.871	0.762	**0.764**	0.717	0.130	0.447	0.059	**0.739**	0.790	**0.747**	**0.747**	0.698
GB-Ens$_{GAT}$	0.857	0.817	**0.827**	0.763	0.150	0.937	0.107	**1.000**	0.850	**0.538**	0.537	0.510
GB-Ens$_{GCN}$	0.835	0.727	**0.730**	0.656	0.540	0.537	0.439	**0.819**	0.770	0.503	**0.504**	0.481

Table 17: OOD detection on Amazon Photos measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.901	0.720	**0.721**	0.694	0.100	**1.000**	0.999	**1.000**	0.800	**0.693**	**0.693**	0.612
CUQ-GAT	0.915	**0.842**	**0.842**	0.823	0.220	0.927	0.927	**0.936**	0.870	**0.510**	**0.510**	0.497
CUQ-GCN	0.912	**0.802**	**0.802**	0.738	0.130	0.948	0.947	**0.953**	0.900	**0.503**	**0.503**	0.474
GPN	0.900	0.835	0.835	**0.892**	0.910	0.522	0.522	**0.568**	0.910	0.552	0.552	**0.565**
GPN-GD$_{EVI}$	0.900	0.812	0.812	**0.855**	0.410	0.712	**0.719**	0.689	0.910	0.563	0.563	**0.573**
GPN-GD$_{LAT}$	0.899	0.810	0.810	**0.855**	0.790	0.561	0.561	**0.583**	0.900	0.554	0.554	**0.579**
LOP-GPN	0.892	0.865	**0.897**	0.569	0.880	0.763	0.683	**0.951**	0.870	0.679	0.584	**0.848**
S-BGCN-K	0.446	0.629	**0.630**	0.359	0.250	0.126	0.071	**0.976**	0.360	0.520	**0.522**	0.477
S-BGCN-T-K	0.717	0.664	**0.665**	0.361	0.210	0.196	0.084	**0.997**	0.640	0.551	**0.553**	0.488
G-Ens$_{APPNP}$	0.930	0.843	**0.844**	0.807	0.135	0.481	0.145	**0.681**	0.900	0.725	**0.726**	0.715
G-Ens$_{GAT}$	0.916	0.786	**0.801**	0.703	0.290	0.895	0.289	**0.999**	0.900	0.497	0.498	**0.520**
G-Ens$_{GCN}$	0.921	**0.847**	**0.847**	0.839	0.240	0.331	0.270	**0.590**	0.900	0.530	**0.531**	0.489
GB-Ens$_{APPNP}$	0.926	0.777	**0.778**	0.740	0.140	0.504	0.164	**0.737**	0.890	0.718	**0.719**	0.701
GB-Ens$_{GAT}$	0.887	0.832	**0.866**	0.810	0.270	0.906	0.251	**1.000**	0.890	0.471	0.474	**0.486**
GB-Ens$_{GCN}$	0.862	0.739	**0.741**	0.497	0.230	0.158	0.137	**0.791**	0.860	0.526	0.527	**0.531**

Table 18: OOD detection on Coauthor CS measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.910	**0.770**	**0.770**	0.741	0.060	**1.000**	**1.000**	**1.000**	0.830	**0.553**	**0.553**	0.507
CUQ-GAT	0.787	0.676	**0.678**	0.639	0.580	**0.663**	0.645	**0.663**	0.820	**0.492**	**0.492**	0.452
CUQ-GCN	0.797	**0.730**	**0.730**	0.658	0.110	**0.979**	0.940	**0.979**	0.770	0.538	**0.541**	0.519
GPN	0.884	0.755	0.755	**0.931**	0.860	0.531	0.531	**0.557**	0.860	0.537	0.537	**0.550**
GPN-GD$_{EVI}$	0.872	0.771	0.760	**0.919**	0.840	0.522	0.522	**0.539**	0.840	0.524	0.524	**0.538**
GPN-GD$_{LAT}$	0.885	0.716	0.716	**0.928**	0.880	**0.541**	**0.541**	0.515	0.880	0.545	0.545	**0.556**
LOP-GPN	0.886	0.706	**0.883**	0.149	0.880	0.678	0.629	**0.723**	0.870	0.620	**0.626**	0.588
S-BGCN-K	0.350	0.659	**0.660**	0.551	0.240	0.148	0.112	**0.924**	0.440	0.465	0.465	**0.517**
S-BGCN-T-K	0.709	0.642	**0.643**	0.543	0.190	0.144	0.090	**0.994**	0.730	**0.542**	**0.542**	0.488
G-Ens$_{APPNP}$	0.934	**0.940**	**0.940**	0.934	0.105	0.724	0.312	**0.889**	0.875	**0.736**	**0.736**	0.713
G-Ens$_{GAT}$	0.915	0.914	**0.920**	0.855	0.155	0.805	0.156	**0.958**	0.870	0.580	**0.583**	0.551
G-Ens$_{GCN}$	0.926	0.924	0.924	**0.939**	0.180	0.558	0.374	**0.788**	0.880	**0.584**	**0.584**	0.566
GB-Ens$_{APPNP}$	0.924	0.826	**0.828**	0.779	0.100	0.750	0.289	**0.963**	0.870	**0.753**	0.752	0.745
GB-Ens$_{GAT}$	0.888	0.728	**0.757**	0.607	0.130	0.872	0.082	**1.000**	0.850	0.570	**0.576**	0.536
GB-Ens$_{GCN}$	0.865	0.657	**0.670**	0.410	0.200	0.191	0.114	**0.810**	0.800	**0.544**	0.543	0.541

Table 19: OOD detection on Coauthor Physics measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.948	**0.861**	**0.861**	0.814	0.240	**1.000**	0.999	**1.000**	0.920	**0.594**	**0.594**	0.544
CUQ-GAT	0.945	**0.848**	**0.848**	0.784	0.260	0.998	0.996	**1.000**	0.930	**0.487**	**0.487**	0.417
CUQ-GCN	0.944	**0.826**	**0.826**	0.782	0.300	**0.992**	**0.992**	**0.992**	0.900	**0.524**	**0.524**	0.474
GPN	0.938	0.876	0.876	**0.957**	0.900	0.562	0.562	**0.607**	0.900	**0.609**	**0.609**	0.600
GPN-GD$_{\text{EVI}}$	0.934	0.912	0.897	**0.957**	0.920	0.613	0.609	**0.643**	0.920	0.618	0.615	**0.641**
GPN-GD$_{\text{LAT}}$	0.926	0.902	0.902	**0.961**	0.920	0.547	0.547	**0.581**	0.920	0.561	0.561	**0.578**
LOP-GPN	0.931	0.832	**0.877**	0.698	0.890	0.751	0.702	**0.767**	0.870	0.701	**0.712**	0.634
S-BGCN-K	0.863	**0.905**	**0.905**	0.814	0.430	0.284	0.197	**0.959**	0.820	**0.593**	**0.593**	0.568
S-BGCN-T-K	0.944	**0.885**	**0.885**	0.826	0.360	0.400	0.290	**0.965**	0.960	**0.601**	**0.601**	0.581
G-Ens$_{\text{APPNP}}$	0.960	**0.940**	**0.940**	0.934	0.210	0.409	0.279	**0.514**	0.960	0.750	**0.751**	0.684
G-Ens$_{\text{GAT}}$	0.953	0.914	**0.920**	0.855	0.270	0.617	0.257	**0.695**	0.960	**0.606**	**0.606**	0.600
G-Ens$_{\text{GCN}}$	0.956	0.924	0.924	**0.939**	0.330	0.574	0.467	**0.730**	0.960	**0.610**	**0.610**	0.592
GB-Ens$_{\text{APPNP}}$	0.954	0.928	**0.931**	0.896	0.210	0.750	0.420	**0.868**	0.900	0.763	**0.764**	0.745
GB-Ens$_{\text{GAT}}$	0.932	0.907	**0.916**	0.844	0.280	0.817	0.307	**0.990**	0.900	**0.600**	**0.600**	0.595
GB-Ens$_{\text{GCN}}$	0.924	0.885	**0.892**	0.733	0.400	0.407	0.362	**0.701**	0.900	**0.604**	**0.604**	0.597

Table 20: OOD detection on OGBN-ArXiv measured by AUROC.

Model	ID ACC	LOC			FI Gaussian				FI Bernoulli			
		TU	AU	EU	OOD ACC	TU	AU	EU	OOD ACC	TU	AU	EU
CUQ-APPNP	0.697	**0.685**	**0.685**	0.662	0.390	**0.823**	**0.823**	0.811	0.680	0.581	0.578	**0.676**
CUQ-GAT	N/A	0.395	**0.405**	0.387	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CUQ-GCN	N/A	**0.523**	0.516	**0.523**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
GPN	0.696	0.706	0.705	**0.747**	0.660	0.540	0.539	**0.604**	0.660	0.540	0.539	**0.604**
GPN-GD$_{\text{EVI}}$	0.675	0.690	0.690	**0.730**	0.680	0.534	0.534	**0.601**	0.680	0.533	0.533	**0.601**
GPN-GD$_{\text{LAT}}$	0.701	0.689	0.689	**0.753**	0.690	0.548	0.548	**0.597**	0.690	0.548	0.548	**0.597**
LOP-GPN	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
S-BGCN-K	0.622	0.647	**0.648**	0.605	0.420	0.540	0.469	**0.775**	0.640	0.591	**0.601**	0.438
S-BGCN-T-K	0.645	0.637	**0.639**	0.590	0.480	0.600	0.499	**0.792**	0.660	0.586	**0.597**	0.450
G-Ens$_{\text{APPNP}}$	0.706	0.642	0.641	**0.708**	0.340	0.540	0.508	**0.942**	0.740	0.598	**0.598**	0.442
G-Ens$_{\text{GAT}}$	0.710	0.695	0.694	**0.702**	0.550	0.651	0.601	**0.888**	0.720	0.529	**0.529**	0.511
G-Ens$_{\text{GCN}}$	0.714	0.667	0.666	**0.705**	0.580	0.596	0.584	**0.680**	0.710	**0.509**	0.509	0.493
GB-Ens$_{\text{APPNP}}$	0.661	0.607	0.589	**0.608**	0.150	0.279	**0.904**	0.364	0.530	0.716	0.614	**0.718**
GB-Ens$_{\text{GAT}}$	0.663	**0.658**	0.551	0.649	0.150	0.232	**0.966**	0.470	0.590	0.555	0.542	**0.558**
GB-Ens$_{\text{GCN}}$	0.644	**0.636**	0.598	0.634	0.380	0.514	**0.701**	0.558	0.610	**0.573**	**0.573**	0.572

Table 21: Computational cost on CoraML.

Model	Params	Models	Preproc. (s)	Epoch (ms)	Fwd	Infer. (ms)	Peak (MiB)
CUQ-APPNP	186,620	1	–	17.0	1	7.0	268
CUQ-GAT	195,196	1	–	16.9	1	6.6	253
CUQ-GCN	194,940	1	–	16.4	1	6.9	235
GPN	186,620	1	–	16.4	1	6.8	173
GPN-GD$_{\text{EVI}}$	186,620	1	–	17.8	1	7.0	202
GPN-GD$_{\text{LAT}}$	186,620	1	–	16.7	1	6.9	188
GPN-LOP	186,620	1	–	24.7	1	15.3	320
S-BGCN-K	46,199	1	0.1	2.4	100	80.0	92
S-BGCN-T-K	46,199	1	0.5	2.7	100	80.0	92
G-Ens$_{\text{APPNP}}$	184,775	10	–	3.5	10	17.0	55
G-Ens$_{\text{GAT}}$	184,917	10	–	3.3	10	13.0	64
G-Ens$_{\text{GCN}}$	184,775	10	–	2.1	10	9.0	177
GB-Ens$_{\text{APPNP}}$	184,775	100	32.5	3.6	100	170.0	167
GB-Ens$_{\text{GAT}}$	184,917	100	32.5	3.3	100	140.0	176
GB-Ens$_{\text{GCN}}$	184,775	100	32.5	2.1	100	100.0	175

Table 22: Computational cost on CoauthorPhysics.

Model	Params	Models	Preproc. (s)	Epoch (ms)	Fwd	Infer. (ms)	Peak (MiB)
CUQ-APPNP	540,564	1	–	46.6	1	24.6	2,886
CUQ-GAT	549,140	1	–	36.4	1	19.0	2,760
CUQ-GCN	548,884	1	–	34.9	1	18.8	2,633
GPN	540,564	1	–	32.2	1	17.1	2,127
GPN-GD$_{\text{EVI}}$	540,564	1	–	32.3	1	16.1	2,127
GPN-GD$_{\text{LAT}}$	540,564	1	–	31.9	1	16.2	2,249
GPN-LOP	540,564	1	–	1835.1	1	1824.7	28,643
S-BGCN-K	538,949	1	18.4	12.0	100	620.0	2,540
S-BGCN-T-K	538,949	1	20.5	11.7	100	620.0	2,545
G-Ens$_{\text{APPNP}}$	2,155,781	10	–	21.0	10	91.0	1,883
G-Ens$_{\text{GAT}}$	1,078,159	10	–	18.6	10	72.0	2,370
G-Ens$_{\text{GCN}}$	2,155,781	10	–	26.0	10	119.0	2,905
GB-Ens$_{\text{APPNP}}$	2,155,781	100	617.5	21.0	100	900.0	1,858
GB-Ens$_{\text{GAT}}$	1,078,159	100	617.5	18.5	100	720.0	2,370
GB-Ens$_{\text{GCN}}$	2,155,781	100	617.5	26.0	100	1180.0	2,905

Table 23: Computational cost on OGBN-ArXiv.

Model	Params	Models	Preproc. (s)	Epoch (ms)	Fwd	Infer. (ms)	Peak (MiB)
CUQ-APPNP	111,152	1	–	778.6	1	505.4	48,500
CUQ-GAT	243,760	1	–	930.0	1	380.3	43,686
CUQ-GCN	242,736	1	–	537.0	1	379.7	38,870
GPN	111,152	1	–	509.5	1	367.4	38,664
GPN-GD$_{\text{EVI}}$	111,152	1	–	575.0	1	367.4	47,412
GPN-GD$_{\text{LAT}}$	111,152	1	–	522.6	1	367.0	42,598
GPN-LOP	OOM	OOM	OOM	OOM	OOM	OOM	OOM
S-BGCN-K	10,856	1	7331.9	22.0	100	770.0	1,594
S-BGCN-T-K	10,856	1	7328.0	21.9	100	780.0	1,620
G-Ens$_{\text{APPNP}}$	110,120	10	–	67.2	10	3040.0	1,024
G-Ens$_{\text{GAT}}$	43,896	10	–	67.3	10	2540.0	5,452
G-Ens$_{\text{GCN}}$	109,096	10	–	80.8	10	3820.0	5,541
GB-Ens$_{\text{APPNP}}$	110,120	100	8362.5	67.3	100	2970.0	1,025
GB-Ens$_{\text{GAT}}$	43,896	100	8362.5	67.1	100	2550.0	5,452
GB-Ens$_{\text{GCN}}$	109,096	100	8362.5	81.3	100	3820.0	5,541