WHO KNOWS WHAT? ONE-SHOT FEDERATED GRAPH LEARNING VIA DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

One-shot federated graph learning is well suited to communication-constrained environments because each client communicates with the server only once. However, this setting is particularly challenging when clients hold graphs that differ in their node features, label distributions, connectivity patterns, and local topology. Conventional aggregation treats each client uniformly across all prediction classes, overlooking variations in class-specific expertise caused by heterogeneous local data and graph structure. We propose **CDRA**, a lightweight **C**lass-wise **D**iffusion **R**eliability **A**ggregation framework for one-shot federated graph learning under severe client heterogeneity. Each client trains a local graph neural network and constructs compact class-conditioned diffusion signatures that capture label-specific structural patterns within its graph. The server uses Hellinger distance to compare these signatures with federation-level class references and estimate each client's reliability for different prediction classes. These reliability estimates, together with client size and class support, determine class-wise aggregation weights for combining local model predictions. Experiments on seven graph benchmarks under severe non-IID heterogeneity show that CDRA improves predictive performance while substantially reducing runtime, peak GPU memory, and upload communication relative to reconstruction-based one-shot federated graph learning methods. Anonymized code is available at: https://anonymous.4open.science/r/CDRA/README.md

1 INTRODUCTION

Federated learning (FL) has been established as a compelling paradigm for training machine learning models across decentralized clients without requiring direct access to raw data. Its advantages are particularly evident in scenarios where data are inherently decentralized and privacy, ownership, or communication limitations make centralized training infeasible (Chung et al., 2026). While substantial progress has been made in federated learning for Euclidean data, extending these ideas to graph-structured data remains considerably more challenging. In federated graph learning, each client may observe a graph with different topology, feature statistics, label distributions, and structural semantics, making the overall federation highly heterogeneous (Tian et al., 2026; Baek et al., 2023). These discrepancies are particularly problematic because graph neural networks are sensitive not only to feature shifts but also to variations in connectivity patterns and local relational structure.

The problem becomes even harder in the *one-shot* federated setting, where each client communicates with the server only once after local training (Yang et al., 2026; Yao et al., 2026; Zhang et al., 2022). This regime is attractive in settings with limited bandwidth and unstable communication, but it removes the iterative correction mechanism that underpins conventional federated optimization. As a result, the server must produce a strong global predictor from a single round of uploaded client information, despite non-IID data, class imbalance, and architecture-level uncertainty. Existing one-shot federated approaches often rely on server-side proxy data, synthetic data generation, or distillation-based surrogates (Wan et al., 2025; Qian et al., 2025; Dai et al., 2024; Kang et al., 2025). However, such strategies can introduce substantial additional overhead at the server, since generating, storing, or optimizing surrogate data or generator-based models increases computational and memory cost and complicates the overall training pipeline.

These limitations motivate a simpler alternative: rather than reconstructing surrogate data or training additional models at the server, clients can directly communicate compact summaries of the structural

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.

information already present in their local graphs. Following this idea, our method represents each locally supported class through a lightweight graph-diffusion signature that describes how it is structurally associated with other classes. The signatures require only a small number of sparse propagation operations, add little communication beyond the local model upload, and eliminate the need for auxiliary generators, public data, or iterative client-server interaction. The server uses these summaries to estimate each client's reliability separately for different prediction classes and accordingly fuse their local model predictions.

Overall, our new ideas and findings in this paper are as follows:

- We propose **CDRA**, a one-shot federated graph-learning framework that captures the structural relationships of each class within a client's local graph using compact class-conditioned diffusion signatures.
- We introduce a Hellinger-margin reliability mechanism that compares each client's class signature with same-class and competing-class federation references, enabling prediction fusion with a distinct client-weight vector for each output class.
- Extensive experiments demonstrate that CDRA provides effective aggregation under severe client heterogeneity while adding minimal computation and communication beyond the one-shot local-model upload.

2 Related Work

In this section, we review the literature most relevant to our study, with a focus on the key research areas that form the foundation of our approach. We specifically discuss prior work in federated learning and related paradigms that motivate our method.

2.1 Federated Learning

FL enables collaborative model training without sharing raw local data. Early methods such as *FedAvg* (McMahan et al., 2017) established the standard client-server paradigm by aggregating locally trained updates at a central server. Subsequent approaches, such as *FedProx* (Li et al., 2020), improve robustness under statistical heterogeneity across clients. Despite their effectiveness, these methods rely on multiple communication rounds and repeated synchronization. This incurs significant overhead, making them less suitable for bandwidth-limited or communication-constrained environments.

2.2 Federated Graph Learning

Extending FL to graph-structured data introduces additional challenges beyond those encountered in standard tabular or vision tasks. In graph settings, clients may differ not only in label distributions and feature statistics, but also in graph topology, neighborhood structure, and local semantic patterns. To address these issues, federated graph learning methods such as *FedPub* (Baek et al., 2023), *FGSSL* (Huang et al., 2023), and *FedTAD* (Zhu et al., 2024) have been proposed to better capture graph-specific information under decentralized settings. These methods have advanced the field by incorporating graph-aware architectures, structural regularization, and knowledge transfer mechanisms. However, many of them still depend on iterative communication and repeated exchange between clients and server, which limits their suitability for one-shot federated scenarios.

2.3 One-Shot Federated Learning

One-shot FL has gained increasing attention as a communication-efficient alternative to traditional multi-round federated optimization. In this setting, each client communicates with the server only once, making global knowledge integration substantially more challenging because iterative refinement is unavailable. Existing methods such as *FedCVAE* (Heinbaugh et al., 2023), *FedSD2C* (Zhang et al., 2024), and *FENS* (Allouah et al., 2024) address this difficulty through synthetic data generation, knowledge distillation, surrogate information, or server-side ensemble learning. Although these approaches reduce client-server communication, their auxiliary generation, reconstruction, or optimization procedures can increase computational complexity and become sensitive to severe

client heterogeneity. More importantly, they are primarily designed for conventional non-graph data and do not explicitly account for graph connectivity or variations in local topology. Consequently, directly applying them to federated graph learning may overlook the structural information required to reconcile clients with heterogeneous graph distributions.

2.4 One-Shot Federated Graph Learning

Recent studies have extended one-shot federated learning to graph-structured data, where knowledge must be integrated from disconnected and heterogeneous client graphs. GHOST (Qian et al., 2025) and OASIS (Wan et al., 2025) address this challenge using client-side graph generators, synthesizers, proxy-graph construction, and server-side knowledge transfer. Although effective, these reconstruction-based pipelines introduce additional training, communication, memory, and computational costs beyond local GNN learning, potentially limiting their use on resource-constrained clients.

This motivates a central question: *Can class-specific structural knowledge be exploited in one-shot federated graph learning (OFGL) without training auxiliary generators or reconstructing proxy graphs?* CDRA addresses this question using compact class-conditioned diffusion signatures computed through sparse local propagation. Clients upload these signatures with their local models and class-support counts, after which the server estimates class-specific reliability and directly fuses local predictions. CDRA thereby avoids surrogate-data generation, graph reconstruction, and auxiliary global-model optimization while retaining a lightweight one-round pipeline.

3 Preliminaries

3.1 One-Shot Federated Graph Learning

We consider a federation of K clients. Each client $k \in \{1, \dots, K\}$ holds a private local graph

$$\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k, \mathbf{X}_k, \mathbf{y}_k), \tag{1}$$

where $\mathcal{V}_k$ and $\mathcal{E}_k$ denote the node and edge sets, $\mathbf{X}_k \in \mathbb{R}^{N_k \times d}$ is the node-feature matrix, and $\mathbf{y}_k \in \{1, \dots, C\}^{N_k}$ contains the node labels. Here, $N_k = |\mathcal{V}_k|$, d denotes the feature dimension, and C is the number of classes. Each client has a labeled training subset $\mathcal{V}_k^{\mathrm{tr}} \subseteq \mathcal{V}_k$. The local graphs may differ in size, label distribution, connectivity, and topology.

Client k independently trains a graph classifier f_{θ_k} using only $\mathcal{G}_k$. For a node $v \in \mathcal{V}_k$, the corresponding predictive distribution is

$$\mathbf{p}_k(v) = \mathrm{softmax}\left(f_{\theta_k}(\mathcal{G}_k, v)\right) \in \Delta^C, \tag{2}$$

In the one-shot setting, clients do not exchange intermediate model updates with the server. Each client communicates only once after local training, after which the server constructs a global prediction rule from the uploaded client information. Raw node features, edges, and node-level labels remain local.

3.2 Graph Propagation

Let $\mathbf{A}_k$ be the adjacency matrix of $\mathcal{G}_k$. After adding self-loops, the normalized propagation operator (Kipf & Welling, 2017) is

$$\mathbf{P}_k = \widetilde{\mathbf{D}}_k^{-1/2} \widetilde{\mathbf{A}}_k \widetilde{\mathbf{D}}_k^{-1/2}, \qquad \widetilde{\mathbf{A}}_k = \mathbf{A}_k + \mathbf{I}, \tag{3}$$

where $\widetilde{\mathbf{D}}_k$ is the degree matrix of $\widetilde{\mathbf{A}}_k$. Applying $\mathbf{P}_k^h$ propagates a node-level signal through paths of up to h steps in the local graph, while the added self-loops preserve each node's own signal during propagation.

3.3 Hellinger Distance

Following the normalized convention of the Hellinger distance (Gibbs & Su, 2002), for two discrete probability distributions $\mathbf{p}, \mathbf{q} \in \Delta^{C-1}$, we define

$$H^2(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{j=1}^{C} \left(\sqrt{p_j} - \sqrt{q_j}\right)^2. \tag{4}$$

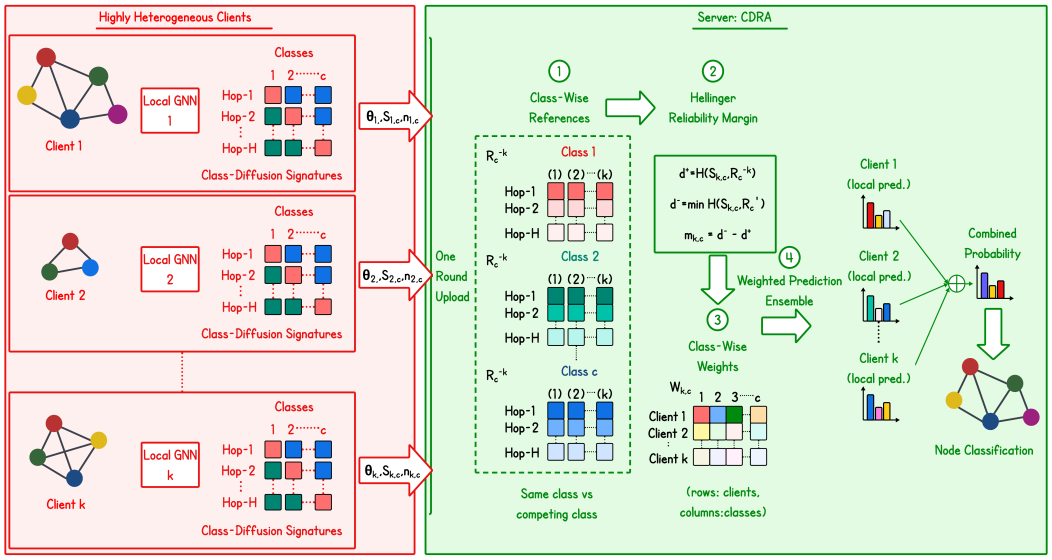


Figure 1: Overview of CDRA. Each client independently trains a local GNN and constructs compact class-conditioned diffusion signatures from its private graph. In a single communication round, clients upload their trained models, eligible class signatures, and class-support counts. The server forms leave-one-client-out class references, measures structural reliability using Hellinger-based separation margins, and combines local predictions through reliability-aware class-wise aggregation. Raw graphs and node-level information remain local.

Hellinger distance is symmetric, bounded, and remains well defined when either distribution contains zero entries. These properties make it suitable for comparing sparse class-mixing profiles arising from highly non-IID client graphs.

4 Methodology

4.1 Overview

Figure 1 presents the proposed **C**lass-wise **D**iffusion **R**eliability **A**ggregation (**CDRA**) framework. CDRA augments each locally trained model with compact signatures that describe the structural behavior of individual classes within the corresponding client graph. The server compares these signatures with federation-level class references, estimates a separate reliability score for each client-class pair, and uses the resulting weights to combine local predictions. The complete procedure requires only one client-to-server communication round and does not involve graph reconstruction, synthetic data generation, auxiliary generators, or additional server-side model training.

4.2 Local Class-Conditioned Diffusion Signatures

Client k constructs a label-indicator matrix

$$\mathbf{Y}_k \in \{0,1\}^{N_k \times C}, \tag{5}$$

where $\mathbf{Y}_k[i,c] = 1$ if $i \in \mathcal{V}_k^{\mathrm{tr}}$ belongs to class c, and $\mathbf{Y}_k[i,c] = 0$ otherwise. Labels outside the local training subset are not used.

Starting from $\mathbf{Y}_k$, the client performs H propagation steps:

$$\mathbf{Z}_k^{(h)} = \mathbf{P}_k^h \mathbf{Y}_k, \qquad h \in \{1, \ldots, H\}. \tag{6}$$

The class-mixing matrix at hop h is then

$$\mathbf{M}_k^{(h)} = \mathrm{RowNorm}\left(\mathbf{Y}_k^\top \mathbf{P}_k^h \mathbf{Y}_k\right) \in \mathbb{R}^{C \times C}. \tag{7}$$

The row $\mathbf{M}_k^{(h)}[c,:]$ summarizes how training nodes of class c are structurally associated with the remaining classes after h propagation steps.

CDRA represents class c on client k using

$$\mathbf{S}_{k,c} = \begin{bmatrix} \mathbf{M}_k^{(1)}[c,:] \\ \vdots \\ \mathbf{M}_k^{(H)}[c,:] \end{bmatrix} \in \mathbb{R}^{H\times C}. \tag{8}$$

The signature is independent of the dimensionality of the node features and contains only (HC) values. Its construction requires sparse propagation with $\mathbf{P}_k$ and small class-level matrix operations.

Let

$$n_{k,c} = \left|\left\{i \in \mathcal{V}_k^{\mathrm{tr}} \,:\, y_{k,i} = c\right\}\right| \tag{9}$$

denote the local support of class c on client k. To avoid transmitting signatures estimated from insufficient local evidence, client k uploads $\mathbf{S}_{k,c}$ only when $n_{k,c} \geq \tau$, where τ denotes the minimum-support threshold.

4.3 One-Shot Client Upload

After local training and signature construction, client k defines the set of eligible classes as

$$\mathcal{C}_k^{\tau} = \{c \in \{1,\dots,C\} : n_{k,c} \geq \tau\}. \tag{10}$$

The client then uploads the one-shot payload

$$\mathcal{U}_k = \left(\theta_k, \{(\mathbf{S}_{k,c}, n_{k,c})\}_{c\in\mathcal{C}_k^{\tau}}\right), \tag{11}$$

where θ_k denotes the parameters of the locally trained model, while $\mathbf{S}_{k,c}$ and $n_{k,c}$ denote the diffusion signature and local support of each eligible class, respectively.

Thus, the client communicates its trained model once, together with the supported class signatures and support counts. If every class satisfies the threshold, the additional structural payload contains HC^2 floating-point values and C counts. No node-level features, edges, embeddings, or labels are transmitted.

4.4 Class-Wise Federation References

The server first constructs a reference signature for each class. Let

$$\alpha_k = \frac{|\mathcal{V}_k^{\mathrm{tr}}|}{\sum_{j=1}^{K}\left|\mathcal{V}_j^{\mathrm{tr}}\right|} \tag{12}$$

denote the normalized client-size prior. The contribution mass of client k for class c is defined as

$$a_{k,c} = \alpha_k n_{k,c}. \tag{13}$$

For each eligible client-class pair (k,c), the server constructs a leave-one-client-out reference using the signatures of the remaining clients:

$$\mathbf{R}_c^{-k} = \frac{\sum\limits_{\substack{j\neq k \\ n_{j,c}\geq\tau}} a_{j,c}\mathbf{S}_{j,c}}{\sum\limits_{\substack{j\neq k \\ n_{j,c}\geq\tau}} a_{j,c}}. \tag{14}$$

Because each row of $\mathbf{S}_{j,c}$ is a probability distribution, their normalized weighted average preserves the same distributional interpretation. Excluding client k prevents its reliability from being estimated through direct agreement with its own signature. The reference $\mathbf{R}_c^{-k}$ is considered available only when at least one other eligible client contributes to class c.

When client k is the only client supporting class c, an independent reference cannot be formed. CDRA treats the corresponding structural evidence as unavailable and assigns a neutral diffusion margin rather than rewarding self-agreement.

4.5 CLASS-SPECIFIC RELIABILITY MARGIN

To compare two diffusion signatures $\mathbf{S}, \mathbf{T} \in \mathbb{R}^{H \times C}$, we average the squared Hellinger distances across their diffusion depths:

$$D_{\text{sig}}(\mathbf{S}, \mathbf{T}) = \sqrt{\frac{1}{H} \sum_{h=1}^{H} H^2(\mathbf{S}[h,:], \mathbf{T}[h,:])}. \tag{15}$$

For an eligible client-class pair (k, c), the positive distance is

$$d_{k,c}^{+} = D_{\text{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_{c}^{-k}\right), \tag{16}$$

which measures agreement with the corresponding leave-one-client-out class reference. The nearest competing-class distance is

$$d_{k,c}^{-} = \min_{\substack{c' \neq c \\ \mathbf{R}_{c'}^{-k} \text{ available}}} D_{\text{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_{c'}^{-k}\right). \tag{17}$$

The class-specific structural reliability margin is then

$$m_{k,c} = d_{k,c}^{-} - d_{k,c}^{+}. \tag{18}$$

A positive margin indicates that the client signature is closer to its corresponding class reference than to the nearest competing reference, whereas a negative margin indicates structural ambiguity or disagreement.

Let Ω denote the set of client-class pairs for which both the positive and negative distances are available. We compute the margin scale as

$$s_m = \max\left\{\text{Median}\left(\{|m_{k,c}| : (k,c) \in \Omega\}\right), \epsilon_m\right\}, \tag{19}$$

where $\epsilon_m > 0$ prevents division by a negligibly small scale. The normalized margin is

$$\widetilde{m}_{k,c} = \begin{cases} m_{k,c}/s_m, & (k,c) \in \Omega, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Thus, unavailable structural evidence produces a neutral diffusion contribution and does not affect the estimation of s_m.

4.6 RELIABILITY-AWARE CLASS-WISE AGGREGATION

CDRA combines client size, local class support, and structural consistency into the raw class-specific reliability score

$$r_{k,c} = \alpha_k^{1/2} n_{k,c}^{\gamma} \exp\left(\frac{\lambda \widetilde{m}_{k,c}}{T_d}\right), \tag{21}$$

where γ controls the class-support prior, λ determines the contribution of the diffusion margin, and $T_d > 0$ is the diffusion temperature. For each class, these scores are normalized across clients:

$$q_{k,c} = \frac{r_{k,c}}{\sum_{j=1}^{K} r_{j,c}}. \tag{22}$$

To obtain a stabilizing client-level reliability estimate, the server computes the support-weighted average

$$\bar{r}_k = \frac{\sum_{c \in \mathcal{C}_k^{\tau}} n_{k,c} r_{k,c}}{\sum_{c \in \mathcal{C}_k^{\tau}} n_{k,c}}. \tag{23}$$

The corresponding global weight is

$$g_k = \frac{\exp(\log(\bar{r}_k + \epsilon)/T_g)}{\sum_{j=1}^{K} \exp(\log(\bar{r}_j + \epsilon)/T_g)}, \tag{24}$$

where $T_g > 0$ is the global temperature and $\epsilon > 0$ ensures numerical stability. The final aggregation weight assigned to client k for class c is

$$w_{k,c} = \beta q_{k,c} + (1 - \beta) g_k, \tag{25}$$

where $\beta \in [0, 1]$ balances class-specific reliability and global stability.

4.7 Prediction Fusion

Given the predictive distribution $\mathbf{p}_k(v)$ produced by client k for node v, CDRA computes the aggregated score for class c as

$$\widehat{p}_c(v) = \sum_{k=1}^{K} w_{k,c} p_{k,c}(v). \tag{26}$$

The final prediction is

$$\widehat{y}(v) = \arg \max_{c \in \{1,\dots,C\}} \widehat{p}_c(v). \tag{27}$$

Unlike client-level ensembling, which uses one client-weight vector for every output class, CDRA allows each client to contribute differently across classes. A theoretical analysis of CDRA is provided in Appendix C.

5 Experimental Setup

This section presents the datasets, baseline methods, and implementation details used to evaluate the proposed method.

5.1 Datasets

We evaluate the proposed framework on seven benchmark graph datasets: *Cora* (McCallum et al., 2000), *CiteSeer* (Giles et al., 1998), *Chameleon* and *Actor* (Pei et al., 2020), *Roman-Empire* (Platonov et al., 2023), *OGBN-Arxiv* (Hu et al., 2020), and *Coauthor-CS* (Shchur et al., 2018). These datasets span citation networks, webpage and actor co-occurrence graphs, a heterophilous word-dependency graph, a co-authorship network, and a large-scale scholarly graph. Collectively, they provide diverse evaluation conditions in terms of graph scale, feature dimensionality, label-space size, connectivity, and homophily. This benchmark suite therefore allows us to assess CDRA under both conventional homophilous conditions and more challenging settings involving heterophily, structural variation, class imbalance, and large graph scale. Detailed dataset descriptions and statistics are provided in Appendix A.

5.2 Baselines

We evaluate the proposed method against a diverse set of baselines drawn from four different groups. As standard FL baselines, we include *FedAvg* (McMahan et al., 2017), *FedProx* (Li et al., 2020), and *FedNova* (Wang et al., 2020). As graph FL (FGL) baselines, we consider *FedPub* (Baek et al., 2023), *FGSSL* (Huang et al., 2023), *FedGTA* (Li et al., 2023), and *FedTAD* (Zhu et al., 2024). We also consider general one-shot federated learning (OFL) methods, including *FedCVAE* (Heinbaugh et al., 2023), *FedSD2C* (Zhang et al., 2024), and *FENS* (Allouah et al., 2024). Finally, we include *GHOST* (Qian et al., 2025) and *OASIS* (Wan et al., 2025), which are specifically designed for one-shot federated graph learning (OFGL) and therefore constitute the closest competitors to our method. This broad comparison enables us to evaluate the proposed approach against standard federated optimization algorithms, graph-specific federated learning methods, general one-shot techniques, and directly related one-shot federated graph learning frameworks.

5.3 Implementation Details

We simulate a one-shot federation of $K = 10$ clients and partition node labels using a Dirichlet distribution with concentration parameter $\alpha_{\mathrm{Dir}} = 0.01$ (Minka, 2000), representing an extreme non-IID setting. For each dataset, nodes are randomly divided into 20% training, 40% validation,

Table 1: Test accuracy (%) on seven benchmark graph datasets under extreme non-IID one-shot federated settings. Results are reported as mean ± standard deviation over multiple seeds, with standard deviations shown in subscript. Best results are in **bold**, and second-best results are underlined.

Method	Cora	CiteSeer	Chameleon	Actor	Roman-Empire	OGBN-Arxiv	Coauthor-CS	Avg.
Traditional FL								
FedAvg[AISTATS17]	$30.41_{\pm 0.4}$	$34.08_{\pm 9.9}$	$20.86_{\pm 0.4}$	$\underline{25.67}_{\pm 5.2}$	$17.12_{\pm 3.7}$	$\underline{19.44}_{\pm 6.6}$	$19.37_{\pm 5.9}$	23.85
FedProx[MLSys20]	$30.78_{\pm 0.5}$	$\underline{35.34}_{\pm 1.4}$	$19.26_{\pm 0.4}$	$23.47_{\pm 3.3}$	$17.42_{\pm 3.0}$	$18.71_{\pm 5.1}$	$22.54_{\pm 0.1}$	$\underline{23.93}$
FedNova[NeurIPS20]	$12.22_{\pm 1.4}$	$15.59_{\pm 2.5}$	$20.42_{\pm 2.4}$	$19.33_{\pm 0.4}$	$7.28_{\pm 3.1}$	$0.81_{\pm 0.3}$	$11.61_{\pm 4.5}$	12.47
Traditional FGL								
FedPub[ICML23]	$30.20_{\pm 0.4}$	$22.71_{\pm 1.4}$	$\underline{23.36}_{\pm 0.6}$	$22.98_{\pm 2.6}$	$8.91_{\pm 4.2}$	$16.60_{\pm 2.8}$	$9.03_{\pm 1.5}$	19.11
FGSSL[IJCAI23]	$30.32_{\pm 0.3}$	$22.29_{\pm 1.7}$	$23.25_{\pm 0.1}$	$21.03_{\pm 3.0}$	$13.73_{\pm 2.7}$	$15.59_{\pm 2.0}$	$18.70_{\pm 5.4}$	20.70
FedGTA[VLDB24]	$11.54_{\pm 1.2}$	$15.17_{\pm 2.4}$	$21.54_{\pm 1.2}$	$19.10_{\pm 0.3}$	$6.98_{\pm 3.2}$	$7.70_{\pm 0.3}$	$6.53_{\pm 3.4}$	12.65
FedTAD[IJCAI24]	$31.42_{\pm 1.4}$	$19.85_{\pm 1.2}$	$22.60_{\pm 0.4}$	$24.04_{\pm 2.6}$	$14.29_{\pm 0.2}$	$12.35_{\pm 6.1}$	$\underline{25.67}_{\pm 2.2}$	21.46
OFL								
FedCVAE[ICLR23]	$6.90_{\pm 0.5}$	$17.71_{\pm 4.2}$	$16.94_{\pm 0.1}$	$20.13_{\pm 3.5}$	$13.60_{\pm 2.8}$	$5.53_{\pm 0.8}$	$7.72_{\pm 3.4}$	12.65
FedSD2C[NeurIPS24]	$16.64_{\pm 1.3}$	$20.30_{\pm 0.7}$	$18.93_{\pm 2.5}$	$21.13_{\pm 2.9}$	$9.62_{\pm 3.6}$	$0.95_{\pm 0.5}$	$16.41_{\pm 8.7}$	14.85
FENS[NeurIPS24]	$30.29_{\pm 0.3}$	$19.35_{\pm 1.0}$	$21.11_{\pm 1.3}$	$24.38_{\pm 2.1}$	$11.29_{\pm 3.8}$	$15.03_{\pm 1.5}$	$22.52_{\pm 0.0}$	20.57
OFGL								
GHOST[ICML25]	$15.27_{\pm 6.5}$	$23.69_{\pm 2.2}$	$21.08_{\pm 1.0}$	$20.58_{\pm 1.4}$	$10.83_{\pm 2.9}$	$12.51_{\pm 7.5}$	$9.11_{\pm 3.3}$	16.15
OASIS[NeurIPS25]	$\underline{35.88}_{\pm 1.2}$	$34.24_{\pm 2.9}$	$20.13_{\pm 0.1}$	$13.60_{\pm 3.3}$	$\underline{18.03}_{\pm 1.9}$	$14.38_{\pm 7.2}$	$12.73_{\pm 12.2}$	21.28
CDRA (Ours)	$\mathbf{45.95}_{\pm 2.3}$	$\mathbf{40.40}_{\pm 3.9}$	$\mathbf{33.48}_{\pm 7.5}$	$\mathbf{28.57}_{\pm 1.4}$	$\mathbf{20.81}_{\pm 3.3}$	$\mathbf{32.50}_{\pm 6.0}$	$\mathbf{31.12}_{\pm 4.1}$	**33.26**

and 40% testing sets. Each client trains a two-layer GCN with a hidden dimension of 128 for 100 epochs. All methods use identical data partitions, model architectures, optimization settings, and a single client-to-server communication round. Further implementation and hyperparameter details are provided in Appendix B.

6 Experimental Results and Discussion

We evaluate CDRA in terms of its overall predictive performance, robustness under heterogeneous federated conditions, and computational and communication efficiency.

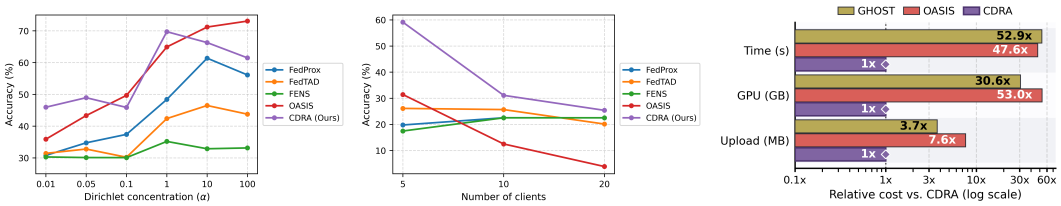
6.1 Overall Performance

Table 1 reports the node-classification performance under the extreme non-IID one-shot setting. CDRA achieves the highest accuracy on all seven datasets and improves the average accuracy from 23.93%, obtained by the strongest baseline, to 33.26%. The gains are particularly pronounced on Cora, Chameleon, and OGBN-Arxiv, where CDRA exceeds the corresponding second-best methods by 10.07, 10.12, and 13.06 percentage points, respectively. Improvements on Actor and Roman-Empire further demonstrate its effectiveness on heterophilous graphs. These results highlight the advantage of estimating client reliability separately for each class rather than assigning a single aggregation weight across all output classes.

The low performance of several baselines reflects the combination of extreme label skew ($\alpha_{\mathrm{Dir}} = 0.01$), missing local class support, graph fragmentation, and the absence of iterative correction. Controlled experiments confirm that representative baselines improve with additional communication or more balanced partitions as shown in Appendix D. Additional results examine CDRA with GAT and GraphSAGE backbones as shown in Appendix F. The ablation study in Appendix E further quantifies the contribution of each component of CDRA. CDRA remains effective in the original setting because its class-conditioned diffusion signatures preserve compact structural evidence, while its support and reliability-aware class-wise aggregation reduces dependence on locally unsupported or structurally inconsistent predictions. Sensitivity to the diffusion depth H is analyzed in Appendix I.

6.2 Robustness and Efficiency

Figure 2 evaluates CDRA under varying degrees of label heterogeneity, different federation sizes, and resource constraints. As shown in Figure 2(a), CDRA remains competitive across the full range of Dirichlet concentrations and performs particularly well in the severely heterogeneous regime targeted by this work. At $\alpha_{\mathrm{Dir}} = 0.01$, CDRA achieves 45.95% accuracy on Cora, outperforming the strongest baseline by 10.07 percentage points. This result indicates that its class-specific reliability estimates remain informative even when local label distributions are highly skewed.



(a) Accuracy versus label heterogeneity on Cora.

(b) Accuracy versus client count on Coauthor-CS.

(c) Relative costs on OGBN-Arxiv.

Figure 2: Robustness and efficiency of CDRA. CDRA remains effective across different label-skew levels and client counts while requiring substantially less runtime, peak GPU memory, and upload communication than reconstruction-based OFGL methods.

Figure 2(b) examines scalability with respect to the number of clients. Although the performance of all methods generally decreases as the graph is divided among more clients, CDRA consistently achieves the highest accuracy for $K \in \{5, 10, 20\}$. Its advantage is especially pronounced with five clients, while its lead is retained as the federation grows. This behavior suggests that class-conditioned diffusion signatures preserve useful structural evidence even when each client observes a progressively smaller portion of the original graph. An extended analysis up to $K = 320$ under a feasible-support configuration ($\alpha_{\mathrm{Dir}} = 5, \tau = 2$) is provided in Appendix G.

Figure 2(c) compares the computational and communication costs of CDRA with the reconstruction-based OFGL methods GHOST and OASIS on OGBN-Arxiv. Relative to CDRA, GHOST and OASIS require approximately $52.9\times$ and $47.6\times$ more execution time, $30.6\times$ and $53.0\times$ more peak GPU memory, and $3.7\times$ and $7.6\times$ greater upload communication, respectively. CDRA avoids client-side graph-generator training, server-side pseudo-graph materialization, and auxiliary global optimization. Its additional communication is limited to compact class-conditioned signatures and support counts. These results demonstrate that CDRA combines robustness to statistical and federation-scale variation with a substantially lighter one-shot aggregation pipeline. The complete per-dataset comparison is provided in Appendix H. CDRA achieves the lowest runtime, peak GPU memory, and upload size on every evaluated dataset, showing that its advantages extend beyond predictive performance to computational and communication efficiency.

7 Conclusion

In this paper, we introduce CDRA, a lightweight OFGL framework designed for highly non-IID clients. CDRA keeps raw graphs, node features, edges, representations, and node-level labels on the clients. Each client communicates only once, uploading its trained model parameters together with compact class-conditioned diffusion signatures and the corresponding class-support counts. The server uses these signatures to estimate the reliability of each client separately for every class through Hellinger-based structural comparisons and fuses the local predictions using reliability-aware class-wise weights. Experiments on seven graph benchmarks show that CDRA consistently outperforms traditional federated learning, federated graph learning, general one-shot federated learning, and one-shot federated graph learning baselines under extreme non-IID conditions. Beyond predictive performance, CDRA substantially reduces runtime, peak GPU memory, and upload communication. These results demonstrate that useful class-specific structural knowledge can be shared and exploited efficiently without transmitting raw graph data or requiring iterative communication. The operating assumptions, threat model, and methodological limitations of CDRA are detailed in Appendices J and K.

AI USE STATEMENT

We used generative AI tools to assist with grammar correction, language refinement, clarity, conciseness, and the presentation of selected passages. These tools were used as writing aids and did not replace the authors' scientific judgment. All AI-assisted text was carefully reviewed and revised by the authors to ensure technical accuracy, consistency with the implemented method and experimental results, appropriate attribution, and originality. Generative AI tools were not used to generate experimental data or determine the reported findings. The authors take full responsibility for the final content of this paper, including all text, claims, analyses, and artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This study uses publicly available benchmark datasets and does not involve new data collection, human-subject experiments, or personally identifiable information. CDRA follows the federated-learning principle of retaining raw node features, edges, representations, and node-level labels on participating clients. The communicated class-conditioned signatures and support counts contain aggregate information and should not be interpreted as providing a formal privacy guarantee. The adversarial assumptions and privacy scope are detailed in Appendix J, while the methodological and deployment limitations are discussed in Appendix K. All methods are evaluated using consistent data partitions, model configurations, and communication constraints, and we report the experimental assumptions to support transparent and reproducible interpretation of the results.

REPRODUCIBILITY STATEMENT

The source code, experiment configurations, and scripts required to reproduce our results are available at `https://anonymous.4open.science/r/CDRA/README.md` and are also included in the supplementary material. The experimental protocol and principal training settings are described in Section 5.3, with complete software, optimization, and CDRA-specific hyperparameter details provided in Appendix B. Dataset descriptions and statistics are reported separately in Appendix A. Together, these resources support reproduction of the reported results under the same one-shot federated graph learning setting.

REFERENCES

Youssef Allouah, Akash Dhasade, Rachid Guerraoui, Nirupam Gupta, Anne-Marie Kermarrec, Rafael Pinot, Rafael Pires, and Rishi Sharma. Revisiting ensembling in one-shot federated learning. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 68500–68527. Curran Associates, Inc., 2024. doi: 10.52202/079017-2188. URL `https://proceedings.neurips.cc/paper_files/paper/2024/file/7ea46207ec9bda974b140fe11d8dd727-Paper-Conference.pdf`.

Jinheon Baek, Wonyong Jeong, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang. Personalized subgraph federated learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 1396–1415. PMLR, 23–29 Jul 2023. URL `https://proceedings.mlr.press/v202/baek23a.html`.

Wu-Chun Chung, Chao-Ai Lo, Yan-Hui Lin, Zhi-Hao Chen, and Che-Lun Hung. Decentralized federated learning with non-iid data: Challenges, trends, and future opportunities. *ACM Comput. Surv.*, 58(8), February 2026. ISSN 0360-0300. doi: 10.1145/3785657. URL `https://doi.org/10.1145/3785657`.

Rong Dai, Yonggang Zhang, Ang Li, Tongliang Liu, Xun Yang, and Bo Han. Enhancing one-shot federated learning through data and ensemble co-boosting. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=tm8s3696Ox`.

Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002. doi: 10.1111/j.1751-5823.2002.tb00178.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2002.tb00178.x.

C. Lee Giles, Kurt D. Bollacker, and Steve Lawrence. CiteSeer: an automatic citation indexing system. In *Proceedings of the ACM International Conference on Digital Libraries*, pp. 89–98. ACM, 1998. Proceedings of the 1998 3rd ACM Conference on Digital Libraries ; Conference date: 23-06-1998 Through 26-06-1998.

Clare Elizabeth Heinbaugh, Emilio Luz-Ricca, and Huajie Shao. Data-free one-shot federated learning under very high statistical heterogeneity. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=_hb4vM3jspB.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 22118–22133. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fb60d411a5c5b72b2e7d3527cfc84fd0-Paper.pdf.

Wenke Huang, Guancheng Wan, Mang Ye, and Bo Du. Federated graph semantic and structural learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI '23, 2023. ISBN 978-1-956792-03-4. doi: 10.24963/ijcai.2023/426. URL https://doi.org/10.24963/ijcai.2023/426.

Myeongkyun Kang, Philip Chikontwe, Soopil Kim, Kyong Hwan Jin, Ehsan Adeli, Kilian M. Pohl, and Sang Hyun Park. Efficient one-shot federated learning on medical data using knowledge distillation with image synthesis and client model adaptation. *Medical Image Analysis*, 105: 103714, 2025. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2025.103714. URL https://www.sciencedirect.com/science/article/pii/S1361841525002610.

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.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In I. Dhillon, D. Papailiopoulos, and V. Sze (eds.), *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020. URL https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf.

Xunkai Li, Zhengyu Wu, Wentao Zhang, Yinlin Zhu, Rong-Hua Li, and Guoren Wang. FedGTA: Topology-aware averaging for federated graph learning. *Proceedings of the VLDB Endowment*, 17 (1):41–50, 2023. ISSN 2150-8097. doi: 10.14778/3617838.3617842.

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. URL https://link.springer.com/article/10.1023/A:1009953814988.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 20–22 Apr 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Thomas Minka. Estimating a Dirichlet distribution, 2000.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1e2agrFvS.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of GNNs under heterophily: Are we really making progress? In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tJbbQfw-5wv.

Jiaru Qian, Guancheng Wan, Wenke Huang, Guibin Zhang, Yuxin Wu, Bo Du, and Mang Ye. GHOST: Generalizable one-shot federated graph learning with proxy-based topology knowledge retention. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=nAk0ENu8LS.

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

Zhihua Tian, Yuan Ding, Rui Zhang, Yao Tang, Jian Liu, and Kui Ren. Toward federated learning of deep graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 38(3): 2028–2039, 2026. doi: 10.1109/TKDE.2026.3652029.

Frank Wan, Jiaru Qian, Wenke Huang, Qilin Xu, Xianda Guo, Boheng Li, Guibin Zhang, Bo Du, and Mang Ye. OASIS: One-shot federated graph learning via wasserstein assisted knowledge integration. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 80038–80068. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/7300dd4d625cb42a438c1b822c77569f-Paper-Conference.pdf.

Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H. Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7611–7623. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/564127c03caab942e503ee6f810f54fd-Paper.pdf.

Xun Yang, Rong Dai, Yonggang Zhang, Ang Li, Tongliang Liu, and Bo Han. Co-Boosting++: Coupled optimization of data and ensemble for one-shot federated learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(7):8269–8282, 2026. doi: 10.1109/TPAMI.2026.3672655.

Dezhong Yao, Tongtong Liu, Yuexin Shi, and Zhiqiang Xu. FedMHO: Heterogeneous one-shot federated learning towards resource-constrained clients. In *Proceedings of the ACM Web Conference 2026*, WWW '26, pp. 5686–5697, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400723070. doi: 10.1145/3774904.3792743. URL https://doi.org/10.1145/3774904.3792743.

Jie Zhang, Chen Chen, Bo Li, Lingjuan Lyu, Shuang Wu, Shouhong Ding, Chunhua Shen, and Chao Wu. DENSE: Data-free one-shot federated learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 21414–21428. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/868f2266086530b2c71006ea1908b14a-Paper-Conference.pdf.

Junyuan Zhang, Songhua Liu, and Xinchao Wang. One-shot federated learning via synthetic distiller-distillate communication. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 102611–102633. Curran Associates, Inc., 2024. doi: 10.52202/079017-3259. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/ba0ad9d1e0c737800b2340b9cd68c208-Paper-Conference.pdf.

Yinlin Zhu, Xunkai Li, Zhengyu Wu, Di Wu, Miao Hu, and Rong-Hua Li. FedTAD: topology-aware data-free knowledge distillation for subgraph federated learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, IJCAI '24, 2024. ISBN 978-1-956792-04-1. doi: 10.24963/ijcai.2024/632. URL https://doi.org/10.24963/ijcai.2024/632.

A Dataset Descriptions

We evaluate CDRA on seven node-classification benchmarks covering citation, webpage, actor co-occurrence, textual, and co-authorship networks. The collection includes graphs with different degrees of homophily, label-space sizes, feature dimensions, and computational scales. Table 2 summarizes their principal statistics. For consistency across datasets and methods, we randomly allocate 20% of the nodes to training, 40% to validation, and 40% to testing.

Table 2: Statistics of the node-classification datasets used in the experiments.

Dataset	#Nodes	#Edges	#Features	#Classes
Cora	2,708	10,556	1,433	7
CiteSeer	3,327	9,104	3,703	6
Chameleon	2,277	36,101	2,325	5
Actor	7,600	30,019	932	5
Roman-Empire	22,662	65,854	300	18
OGBN-Arxiv	169,343	1,166,243	128	40
Coauthor-CS	18,333	163,788	6,805	15

Citation graphs. **Cora** (McCallum et al., 2000) and **CiteSeer** (Giles et al., 1998) represent scientific publications connected through citation relations. Their node attributes are sparse document representations, and their labels identify research areas. Although both are comparatively small citation graphs, their different connectivity and class distributions provide distinct federated partitioning conditions.

Web and actor networks. **Chameleon** and **Actor** (Pei et al., 2020) represent settings in which neighboring nodes do not necessarily share the same label. Chameleon is a Wikipedia-page network whose edges correspond to hyperlinks and whose node labels represent page categories. Actor is an actor-only subgraph extracted from a film-related network: nodes denote actors, edges indicate co-occurrence on Wikipedia pages, and node features encode descriptive keywords. These datasets are included to test CDRA when local connectivity provides weaker homophilous signals.

Roman-Empire. **Roman-Empire** (Platonov et al., 2023) is a heterophilous text-derived graph constructed from the English Wikipedia article on the Roman Empire. Each node represents a word occurrence, while edges encode sequential or syntactic relationships between words. Node features are word embeddings, and labels indicate syntactic roles. Its relatively large label space and heterophilous structure make it a challenging test of class-conditioned structural aggregation.

Large citation graph. **OGBN-Arxiv** (Hu et al., 2020) is a large directed citation network from the Open Graph Benchmark. Its nodes correspond to computer science papers, directed edges represent citations, and node features are derived from the papers' titles and abstracts. Each paper is assigned to one of 40 subject categories. The dataset is substantially larger than the other benchmarks and is used to examine whether CDRA remains effective as graph size and label-space complexity increase.

Co-authorship graph. **Coauthor-CS** (Shchur et al., 2018) models scientific collaboration in computer science. Nodes correspond to researchers, edges connect coauthors, and node attributes are obtained from the keywords of each author's publications. Labels represent research fields. This graph complements the citation benchmarks by introducing a different scholarly relation and a higher-dimensional feature space.

Collectively, these datasets evaluate CDRA across homophilous and heterophilous graphs, small and large label spaces, and graph sizes ranging from a few thousand to more than one hundred thousand nodes. They therefore provide varied conditions for studying class-specific reliability under extreme non-IID client partitions.

B Extended Implementation Details

All experiments are implemented using Python 3.11.7, PyTorch 2.5.1, PyTorch Geometric 2.7.0, NumPy 1.26.4, and scikit-learn 1.7.1. We simulate a federation of $K = 10$ clients and partition the

node labels according to a Dirichlet distribution with concentration parameter $\alpha_{\mathrm{Dir}} = 0.01$ (Minka, 2000). This setting produces highly skewed client-level label distributions, representing an extreme non-IID scenario in which some clients may have limited support for particular classes. For each dataset, nodes are randomly divided into 20% training, 40% validation, and 40% testing sets.

Each client employs a two-layer GCN with hidden dimension 128 and dropout rate 0.3. All local models use the same architecture and are initialized from an identical set of model parameters. Each model is trained for 100 epochs using the Adam optimizer with a learning rate of 5×10^{-3} and weight decay of 4×10^{-4}. The checkpoint with the highest validation accuracy is retained.

For CDRA, we use $H = 3$ diffusion depths and a minimum class-support threshold of $\tau = 10$. The class-support exponent and structural reliability coefficient are set to $\gamma = 0.5$ and $\lambda = 1.0$, respectively. Both the diffusion and global temperatures are set to $T_d = T_g = 1.5$, while the class-global interpolation coefficient is set to $\beta = 0.8$. Class reference signatures are constructed using the leave-one-client-out strategy. The reliability margins are normalized using the median absolute margin over the client-class pairs for which both positive and negative reference distances are available.

During the single communication round, each client uploads its trained model together with the eligible class-conditioned diffusion signatures $\{\mathbf{S}_{k,c}\}_{c\in\mathcal{C}_k^\tau}$ and their corresponding class-support counts $\{n_{k,c}\}_{c\in\mathcal{C}_k^\tau}$. No subsequent client-server communication is performed.

For a controlled comparison, all methods use the same client partitions, data splits, model architecture, initialization, optimization settings, and non-IID configuration. Every baseline is restricted to the same one-shot protocol, comprising a single client-to-server communication stage without iterative refinement. The resulting performance differences therefore primarily reflect the respective aggregation and knowledge-transfer mechanisms rather than differences in communication budgets or experimental conditions.

C Theoretical Analysis

This section establishes three properties of CDRA. First, its signature distance is a bounded metric and remains well defined for sparse distributions. Second, the estimation error of a class-conditioned diffusion profile decreases with its local class support. Third, the sign of the CDRA reliability margin correctly identifies a structurally consistent client whenever the estimation error is smaller than the separation between the population-level class profiles.

C.1 Well-Posedness of the Signature Distance

Recall that the normalized Hellinger distance between $\mathbf{p}, \mathbf{q} \in \Delta^{C-1}$ is

$$H(\mathbf{p}, \mathbf{q}) = \left[\frac{1}{2}\sum_{r=1}^{C}\left(\sqrt{p_r} - \sqrt{q_r}\right)^2\right]^{1/2}. \tag{28}$$

For two diffusion signatures $\mathbf{S}, \mathbf{T} \in (\Delta^{C-1})^H$, CDRA uses

$$D_{\mathrm{sig}}(\mathbf{S}, \mathbf{T}) = \left[\frac{1}{H}\sum_{h=1}^{H} H^2(\mathbf{S}[h,:], \mathbf{T}[h,:])\right]^{1/2}. \tag{29}$$

Lemma 1 (Bounded signature metric)**.** *D_{sig} is a metric on $(\Delta^{C-1})^H$ and satisfies*

$$0 \le D_{\mathrm{sig}}(\mathbf{S}, \mathbf{T}) \le 1. \tag{30}$$

Consequently, whenever the required references are available, $d^+_{k,c}$, $d^-_{k,c}$, and $m_{k,c}$ are well defined and satisfy

$$0 \le d^+_{k,c}, d^-_{k,c} \le 1, \qquad -1 \le m_{k,c} \le 1. \tag{31}$$

Proof. The normalized Hellinger distance is a metric on the probability simplex and is bounded by one. Non-negativity, symmetry, and identity of indiscernibles therefore hold componentwise for every diffusion depth. Moreover, the vector

$$(H(\mathbf{S}[1,:], \mathbf{T}[1,:]), \ldots, H(\mathbf{S}[H,:], \mathbf{T}[H,:]))$$

is aggregated using a scaled Euclidean norm. The triangle inequality for D_{sig} consequently follows from the componentwise triangle inequality for H and Minkowski's inequality. Since every component is at most one, their root-mean-square is also at most one. The bounds on $d^+_{k,c}$, $d^-_{k,c}$, and $m_{k,c} = d^-_{k,c} - d^+_{k,c}$ follow immediately. □

Unlike divergences involving logarithms or division by probability masses, Hellinger distance remains finite when either argument contains zero-valued entries. It is therefore suitable for the sparse class-mixing distributions produced by extreme non-IID partitions.

C.2 Finite-Support Accuracy of Diffusion Profiles

We next relate the accuracy of a local signature to its class support. Fix client k, class c, and diffusion depth h. For a class-c training node i, define its nonnegative propagated class-mass vector as

$$\mathbf{X}^{(h)}_{k,c,i}[r] = \sum_{\substack{j \in \mathcal{V}^{\mathrm{tr}}_k \\ y_{k,j}=r}} [\mathbf{P}^h_k]_{ij}, \qquad r \in \{1, \dots, C\}. \tag{32}$$

The corresponding row of the empirical class-mixing matrix can be written as

$$\mathbf{M}^{(h)}_k[c,:] = \frac{\sum_{i:y_{k,i}=c} \mathbf{X}^{(h)}_{k,c,i}}{\sum_{i:y_{k,i}=c} \|\mathbf{X}^{(h)}_{k,c,i}\|_1}. \tag{33}$$

For the analysis, suppose that the vectors $\{\mathbf{X}^{(h)}_{k,c,i} : y_{k,i} = c\}$ are conditionally independent samples from a client–class distribution. Assume $\|\mathbf{X}^{(h)}_{k,c,i}\|_\infty \le B_h$ and $\|\mathbf{X}^{(h)}_{k,c,i}\|_1 \le L_h$ almost surely, and let

$$b^{(h)}_{k,c} = \mathbb{E}\left[\|\mathbf{X}^{(h)}_{k,c,i}\|_1\right] > 0. \tag{34}$$

Define the population diffusion profile

$$\mathbf{M}^{(h)}_{k,*}[c,:] = \frac{\mathbb{E}[\mathbf{X}^{(h)}_{k,c,i}]}{b^{(h)}_{k,c}}. \tag{35}$$

Proposition 2 (Concentration with local class support)**.** *Let $n = n_{k,c}$. For any $\delta \in (0,1)$, provided that*

$$L_h \sqrt{\frac{\log(4/\delta)}{2n}} \le \frac{b^{(h)}_{k,c}}{2}, \tag{36}$$

the following bound holds with probability at least $1-\delta$:

$$\begin{aligned} &\left\| \mathbf{M}^{(h)}_k[c,:] - \mathbf{M}^{(h)}_{k,*}[c,:] \right\|_1 \\ &\quad \le \frac{2}{b^{(h)}_{k,c}} \left[C B_h \sqrt{\frac{\log(4C/\delta)}{2n}} + L_h \sqrt{\frac{\log(4/\delta)}{2n}} \right]. \end{aligned} \tag{37}$$

Thus, for fixed C, B_h, L_h, and $b^{(h)}_{k,c}$, the profile estimation error is $\mathcal{O}\left(n^{-1/2}_{k,c}\right)$.

Proof. Apply Hoeffding's inequality to each coordinate of the empirical mean $n^{-1}\sum_i \mathbf{X}^{(h)}_{k,c,i}$ and use a union bound over the C coordinates. A second application of Hoeffding's inequality controls the empirical normalization mass $n^{-1}\sum_i \|\mathbf{X}^{(h)}_{k,c,i}\|_1$. Under Equation 36, this empirical denominator is at least $b^{(h)}_{k,c}/2$. Decomposing the difference between the two normalized vectors into numerator and denominator errors gives Equation 37. □

Applying Proposition 2 over all H diffusion depths with a union bound shows that the complete signature $\mathbf{S}_{k,c}$ converges to its population counterpart $\mathbf{S}^*_{k,c}$ as $n_{k,c}$ increases. In particular, the bound

provides a statistical interpretation of the minimum-support rule: τ excludes client–class signatures for which a desired finite-sample accuracy cannot be guaranteed.

The bound also motivates the support factor $n_{k,c}^{\gamma}$ in Equation 21. Because the estimation error scales as $n_{k,c}^{-1/2}$, setting $\gamma = 1/2$ assigns reliability in proportion to the inverse order of the estimator's standard error. Hence, the support term favors signatures supported by stronger local evidence without allowing raw class frequency to dominate the structural reliability margin.

Relation to Hellinger error. For arbitrary probability distributions,

$$H^2(\mathbf{p}, \mathbf{q}) \leq \frac{1}{2}\|\mathbf{p} - \mathbf{q}\|_1. \tag{38}$$

Therefore, Proposition 2 also controls the squared Hellinger error without requiring strictly positive probability entries. If the population profile is bounded away from the simplex boundary, i.e., $\mathbf{M}_{k,*}^{(h)}[c, r] \geq \kappa > 0$ for every r, then the square-root map is locally Lipschitz and the Hellinger error itself decreases as $\mathcal{O}(n_{k,c}^{-1/2})$.

C.3 Correctness of the Reliability Margin

Let $\mathbf{R}_c^* \in (\Delta^{C-1})^H$ denote the population-level structural reference for class c. Define its separation from the nearest competing class by

$$\Delta_c = \min_{c' \neq c} D_{\text{sig}}\left(\mathbf{R}_c^*, \mathbf{R}_{c'}^*\right). \tag{39}$$

A positive Δ_c means that class c has a structurally distinguishable population diffusion profile.

For an eligible client–class pair (k, c), define the local deviation

$$\varepsilon_{k,c} = D_{\text{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_c^*\right) \tag{40}$$

and the leave-one-out reference errors

$$\rho_{k,c} = D_{\text{sig}}\left(\mathbf{R}_c^{-k}, \mathbf{R}_c^*\right), \qquad \rho_{k,-c} = \max_{c' \neq c} D_{\text{sig}}\left(\mathbf{R}_{c'}^{-k}, \mathbf{R}_{c'}^*\right), \tag{41}$$

where the maximum is taken over the available competing references.

Theorem 3 (Positive-margin certificate)**.** *For every eligible pair (k, c) with valid positive and negative references, the CDRA reliability margin satisfies*

$$m_{k,c} \geq \Delta_c - 2\varepsilon_{k,c} - \rho_{k,c} - \rho_{k,-c}. \tag{42}$$

Consequently,

$$2\varepsilon_{k,c} + \rho_{k,c} + \rho_{k,-c} < \Delta_c \quad \Longrightarrow \quad m_{k,c} > 0. \tag{43}$$

Proof. By the triangle inequality,

$$d_{k,c}^{+} = D_{\text{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_c^{-k}\right) \leq \varepsilon_{k,c} + \rho_{k,c}. \tag{44}$$

For any competing class $c' \neq c$, the reverse triangle inequality gives

$$D_{\text{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_{c'}^{-k}\right) \geq D_{\text{sig}}\left(\mathbf{R}_c^*, \mathbf{R}_{c'}^*\right) - \varepsilon_{k,c} - \rho_{k,c'} \tag{45}$$

$$\geq \Delta_c - \varepsilon_{k,c} - \rho_{k,-c}. \tag{46}$$

Taking the minimum over $c' \neq c$ yields $d_{k,c}^{-} \geq \Delta_c - \varepsilon_{k,c} - \rho_{k,-c}$. Subtracting the upper bound on $d_{k,c}^{+}$ proves Equation 42. □

Theorem 3 provides a direct interpretation of the CDRA margin. A positive margin is guaranteed when the structural separation between class c and its nearest competitor exceeds the combined local-signature and reference-estimation errors. The condition also identifies two regimes in which structural reliability is inherently difficult to estimate: classes with weak population separation and client–class pairs with limited local support.

Corollary 4 (Asymptotic margin consistency)**.** *Suppose that class c has positive population-level structural separation, i.e., $\Delta_c > 0$. Assume that, as the number and effective contribution mass of eligible clients increase, the leave-one-client-out reference errors satisfy*

$$\rho_{k,c} \xrightarrow{p} 0, \qquad \rho_{k,-c} \xrightarrow{p} 0. \tag{47}$$

Further suppose that the local class support $n_{k,c}$ increases and that client k is structurally consistent with the population-level class-c profile, so that

$$\varepsilon_{k,c} = D_{\mathrm{sig}}\left(\mathbf{S}_{k,c}, \mathbf{R}_c^*\right) \xrightarrow{p} 0. \tag{48}$$

Then

$$\Pr(m_{k,c} > 0) \longrightarrow 1. \tag{49}$$

Proof. By assumption,

$$2\varepsilon_{k,c} + \rho_{k,c} + \rho_{k,-c} \xrightarrow{p} 0. \tag{50}$$

Since $\Delta_c > 0$, it follows that

$$\Pr(2\varepsilon_{k,c} + \rho_{k,c} + \rho_{k,-c} < \Delta_c) \longrightarrow 1. \tag{51}$$

Whenever this inequality holds, Theorem 3 guarantees that $m_{k,c} > 0$. Therefore,

$$\Pr(m_{k,c} > 0) \longrightarrow 1. \tag{52}$$

□

This corollary shows that, for a structurally consistent client-class pair, the CDRA margin identifies the correct class reference with probability approaching one as both the local signature and the federation-level references become increasingly accurate. The result also explains the role of leave-one-client-out construction: excluding client k prevents $\mathbf{S}_{k,c}$ from contributing to the reference against which it is evaluated. Consequently, the margin measures agreement with structural evidence contributed by other eligible clients rather than agreement induced by self-inclusion.

C.4 Validity of the Final Aggregation Weights

Proposition 5 (Valid class-wise aggregation)**.** *Suppose $r_{k,c} \geq 0$ and $\bar{r}_k \geq 0$, with at least one positive score in every normalization. Then $\mathbf{q}_c = (q_{1,c}, \ldots, q_{K,c})$ and $\mathbf{g} = (g_1, \ldots, g_K)$ belong to the probability simplex. For every $\beta \in [0, 1]$, the final CDRA weight vector $\mathbf{w}_c = (w_{1,c}, \ldots, w_{K,c})$ also belongs to the probability simplex:*

$$w_{k,c} \geq 0, \qquad \sum_{k=1}^{K} w_{k,c} = 1. \tag{53}$$

Proof. The normalized scores $q_{k,c}$ and softmax weights g_k are nonnegative and each sum to one across clients. Since

$$w_{k,c} = \beta q_{k,c} + (1 - \beta) g_k \tag{54}$$

is a convex combination of these two distributions,

$$\sum_{k=1}^{K} w_{k,c} = \beta \sum_{k=1}^{K} q_{k,c} + (1 - \beta) \sum_{k=1}^{K} g_k = 1. \tag{55}$$

Non-negativity follows from $\beta \in [0, 1]$. Any positive floor followed by renormalization preserves these properties. □

Accordingly, for each output class c, CDRA performs a valid convex aggregation of the client predictions. Unlike a conventional ensemble, however, the relevant probability vector $\mathbf{w}_c$ may differ across output classes.

D Controlled Analysis of the One-Shot Evaluation Regime

Several baselines achieve comparatively low accuracy on OGBN-Arxiv under the extreme non-IID one-shot protocol. To determine whether this behavior arises from the evaluation regime rather than an implementation failure, we select FedNova as a representative baseline and conduct a controlled experiment varying the communication budget and degree of label heterogeneity. FedNova is suitable for this analysis because it was originally designed for iterative federated optimization, allowing us to directly examine the effect of restricting a multi-round method to one-shot communication. The one-shot configuration uses one communication round with 100 local epochs, whereas the iterative configurations use 50 rounds with two local epochs per round. All configurations therefore perform 100 local epochs in total but differ in their opportunities for client-server correction.

Table 3: Controlled evaluation of FedNova on OGBN-Arxiv under different communication budgets and Dirichlet label-partition parameters. All settings use the same total number of local training epochs.

Protocol	α_{Dir}	**Rounds**	**Epochs/round**	**Total epochs**	**Accuracy (%)**
One-shot	0.01	1	100	100	$0.81_{\pm 0.3}$
Multi-round	0.01	50	2	100	$24.11_{\pm 7.3}$
Multi-round	0.05	50	2	100	$25.14_{\pm 8.7}$
Multi-round	1	50	2	100	$42.66_{\pm 1.1}$
Multi-round	100	50	2	100	$50.26_{\pm 0.2}$

As shown in Table 3, increasing the communication budget from one to 50 rounds while retaining the extreme non-IID setting $\alpha_{\mathrm{Dir}} = 0.01$ improves accuracy from 0.81% to 24.11%, a gain of 23.30 percentage points. Because both configurations use the same total number of local epochs, this comparison demonstrates the importance of repeated client-server correction for a method designed for iterative federated optimization. Under the 50-round protocol, reducing label heterogeneity produces further improvements. Increasing α_{Dir} from 0.01 to 0.05 yields a modest increase from 24.11% to 25.14%. A larger improvement occurs at $\alpha_{\mathrm{Dir}} = 1$, where accuracy reaches 42.66%, while the near-IID configuration $\alpha_{\mathrm{Dir}} = 100$ achieves 50.26%. The decreasing standard deviation at larger Dirichlet concentrations also indicates that more balanced client partitions produce more stable performance across random seeds. These results confirm that the selected baseline is functional and that its low one-shot accuracy is primarily associated with the combined effects of extreme statistical heterogeneity and the absence of iterative communication. More broadly, the analysis explains why methods developed for conventional multi-round federated optimization may perform poorly when restricted to the deliberately challenging one-shot regime used in our main evaluation.

E Ablation Study

Table 4: Accuracy (%) ablation study of CDRA on Cora and OGBN-Arxiv. Results are reported as mean ± standard deviation over multiple seeds. Best results are shown in **bold**, and second-best results are underlined. The symbol "–" indicates that LOO is not applicable.

Method	**Diffusion**	**Support**	**Size**	**Class-wise**	**LOO**	**Cora**	**OGBN-Arxiv**
Uniform Ensemble	✗	✗	✗	✗	–	$31.42_{\pm 5.3}$	$30.54_{\pm 6.8}$
Size Only	✗	✗	✓	✗	–	$33.59_{\pm 4.8}$	$29.60_{\pm 7.1}$
Support Only	✗	✓	✗	✓	–	$42.50_{\pm 4.5}$	$27.77_{\pm 1.5}$
Diffusion Only	✓	✗	✗	✓	✓	$43.08_{\pm 7.3}$	$28.08_{\pm 1.1}$
Global CDRA	✓	✓	✓	✗	✓	$30.90_{\pm 0.5}$	$30.49_{\pm 7.1}$
CDRA w/o LOO	✓	✓	✓	✓	✗	$\mathbf{46.57_{\pm 0.7}}$	$\underline{31.61_{\pm 6.7}}$
Full CDRA	✓	✓	✓	✓	✓	$\underline{45.95_{\pm 2.3}}$	$\mathbf{32.50_{\pm 6.0}}$

Table 4 evaluates the contribution of the main components of CDRA on Cora and OGBN-Arxiv. Uniform ensembling provides the reference configuration, while the remaining variants progressively

introduce client-size information, class support, diffusion-based reliability, class-wise weighting, and leave-one-client-out (LOO) references.

On Cora, both class support and diffusion information provide substantial improvements over uniform ensembling. Support-only and diffusion-only aggregation achieve 42.50% and 43.08% accuracy, respectively, compared with 31.42% for the uniform ensemble. In contrast, size-only weighting provides a smaller improvement, reaching 33.59%. These results indicate that class-dependent evidence is more informative than assigning clients a single weight based only on their overall dataset size.

The importance of class-wise aggregation is further demonstrated by Global CDRA. Although this variant retains diffusion, support, size, and LOO information, it collapses the resulting reliability estimates into one client-level weight and obtains only 30.90% on Cora. Restoring class-specific weighting increases accuracy to 45.95% for Full CDRA. This result supports the central motivation of CDRA: a client may be reliable for some classes but not for others, particularly under severely imbalanced label partitions.

The effect of LOO references is dataset dependent. On OGBN-Arxiv, Full CDRA improves upon CDRA without LOO from 31.61% to 32.50%, indicating that excluding the evaluated client from its reference reduces the effect of self-agreement. On Cora, CDRA without LOO achieves a slightly higher mean accuracy of 46.57% than Full CDRA's 45.95%, although the difference is small relative to the reported variation. This suggests that LOO provides a principled safeguard against self-reinforcement but does not necessarily improve the mean accuracy in every dataset and random partition.

Overall, no isolated component consistently matches the complete method across both datasets. Full CDRA obtains the strongest result on OGBN-Arxiv and remains competitive with the best ablation on Cora, showing that support, structural reliability, client size, and class-wise aggregation provide complementary signals under different graph and label-distribution conditions.

F Effect of Backbone Choice

Table 5: Test accuracy (%) under different backbone architectures on Cora and Chameleon. Results are reported as mean ± standard deviation over multiple seeds, with standard deviations shown in subscript. Best results in each row are in **bold**, and second-best results are underlined.

Dataset	Backbone	FedProx	FedTAD	FENS	OASIS	CDRA (Ours)
Cora	GAT	$30.26_{\pm 0.3}$	$30.17_{\pm 0.1}$	$31.07_{\pm 0.1}$	$\underline{37.77_{\pm 2.6}}$	$\mathbf{39.24_{\pm 8.6}}$
	GraphSAGE	$30.11_{\pm 0.1}$	$27.09_{\pm 4.4}$	$30.11_{\pm 0.1}$	$\underline{32.00_{\pm 4.2}}$	$\mathbf{36.22_{\pm 5.7}}$
Chameleon	GAT	$\underline{23.10_{\pm 0.4}}$	$21.76_{\pm 1.5}$	$21.83_{\pm 1.4}$	$20.06_{\pm 0.1}$	$\mathbf{32.72_{\pm 2.4}}$
	GraphSAGE	$\mathbf{23.79_{\pm 2.6}}$	$\underline{21.91_{\pm 1.4}}$	$20.89_{\pm 1.3}$	$20.09_{\pm 0.1}$	$21.14_{\pm 1.2}$

Table 5 examines whether CDRA remains effective when the default GCN backbone is replaced by GAT or GraphSAGE. CDRA achieves the highest accuracy in three of the four dataset-backbone configurations. On Cora, it outperforms the second-best method by 1.47 percentage points with GAT and 4.22 percentage points with GraphSAGE. The largest improvement is observed on Chameleon with GAT, where CDRA achieves 32.72% accuracy, exceeding the strongest baseline by 9.62 percentage points.

The results demonstrate that CDRA is not tied to a particular GNN architecture. Its class-conditioned diffusion signatures are computed directly from the local propagation structure and training labels, independently of the hidden representations produced by the local model. The resulting reliability weights can therefore be applied to the class predictions of different GNN backbones without modifying the underlying aggregation procedure.

An exception occurs with GraphSAGE on Chameleon, where CDRA obtains 21.14% accuracy compared with 23.79% for FedProx. This result indicates that the effectiveness of reliability-aware prediction fusion still depends on the quality and structural suitability of the underlying local predictors. In particular, the interaction between neighborhood sampling-based aggregation and

the heterophilous structure of Chameleon may produce less discriminative local class predictions. Overall, however, CDRA's leading performance in three of the four configurations shows that its benefits generally transfer beyond the default GCN backbone.

G CLIENT-SCALING BOUNDARY

The practical client-scaling boundary of CDRA is determined not by a universal number of clients, but by the class support available after a fixed graph is partitioned. Let $N^{\mathrm{tr}} = |\mathcal{V}^{\mathrm{tr}}|$ denote the total number of training nodes. When these nodes are distributed across K clients, the average client training size decreases approximately as N^{tr}/K. The class support $n_{k,c}$ consequently becomes smaller as K increases, particularly under a highly concentrated Dirichlet partition.

This effect is directly relevant to CDRA because client k uploads the signature $\mathbf{S}_{k,c}$ only when $n_{k,c} \geq \tau$. Furthermore, a leave-one-client-out reference for class c is available only when at least two clients satisfy this condition. Therefore, increasing K, decreasing α_{Dir}, and increasing τ jointly reduce the amount of structural evidence available to the server.

A simple finite-support relationship makes this dependence explicit. Because the client-class supports sum to the total number of training nodes,

$$\sum_{k=1}^{K}\sum_{c=1}^{C} n_{k,c} = N^{\mathrm{tr}}, \tag{56}$$

the total number of eligible client-class pairs is bounded by

$$\sum_{k=1}^{K}\sum_{c=1}^{C} \mathbb{I}[n_{k,c} \geq \tau] \leq \frac{N^{\mathrm{tr}}}{\tau}. \tag{57}$$

Thus, for a fixed graph and threshold, the fraction of eligible client-class pairs can decrease at least proportionally to $1/K$ as the federation grows.

Our main experiments use $K = 10$ clients under the deliberately challenging configuration $\alpha_{\mathrm{Dir}} = 0.01$ and $\tau = 10$. Under this configuration, large values of K can leave many classes without sufficient support to construct independent leave-one-out references. To distinguish this finite-data coverage constraint from the computational scalability of CDRA, we conduct an additional scaling experiment using a more balanced partition, $\alpha_{\mathrm{Dir}} = 5$, and a lower fixed minimum-support threshold, $\tau = 2$. These values are retained across all client counts in the extended experiment.

For each setting, we report both predictive accuracy and two structural coverage measures. Eligible-pair coverage is defined as

$$\mathrm{Cov}_{\mathrm{sig}} = \frac{1}{KC}\sum_{k=1}^{K}\sum_{c=1}^{C} \mathbb{I}[n_{k,c} \geq \tau]. \tag{58}$$

Let

$$L_c = \sum_{k=1}^{K} \mathbb{I}[n_{k,c} \geq \tau] \tag{59}$$

denote the number of clients uploading a valid signature for class c. A leave-one-client-out reference is available for class c only when $L_c \geq 2$. Accordingly, the LOO class coverage is

$$\mathrm{Cov}_{\mathrm{LOO}} = \frac{1}{C}\sum_{c=1}^{C} \mathbb{I}[L_c \geq 2]. \tag{60}$$

Here, $\mathrm{Cov}_{\mathrm{sig}}$ denotes the fraction of client-class pairs that produce an eligible signature, while L_c is the number of eligible clients for class c. Consequently, $\mathrm{Cov}_{\mathrm{LOO}}$ measures the fraction of classes supported by at least two eligible clients and therefore admitting an independent leave-one-client-out reference.

As shown in Table 6, CDRA achieves 76.45% accuracy with five clients and 67.13% with 10 clients. Accuracy decreases to 47.32% at $K = 20$, followed by a partial recovery at $K = 40$ and $K = 80$,

Table 6: Extended client-scaling results under a feasible-support partitioning configuration. The Dirichlet concentration and minimum-support threshold are fixed across all client counts.

K	CDRA (Ours)
5	$76.45_{\pm 0.4}$
10	$67.13_{\pm 4.8}$
20	$47.32_{\pm 0.9}$
40	$51.76_{\pm 5.7}$
80	$55.97_{\pm 7.0}$
160	$47.83_{\pm 6.8}$
320	$36.45_{\pm 2.6}$

where it reaches 51.76% and 55.97%, respectively. This non-monotonic behavior reflects variation in the class support and structural evidence retained by different Dirichlet partitions rather than a strictly monotonic dependence on K. At larger federation sizes, accuracy decreases to 47.83% with 160 clients and 36.45% with 320 clients. Thus, increasing the federation size from 10 to 320 clients, a $32\times$ increase, reduces accuracy by 30.68 percentage points.

The overall decline at large K is consistent with the finite-support behavior of CDRA. As the fixed graph is divided among more clients, each client observes fewer labeled nodes, fewer client-class pairs satisfy $n_{k,c} \geq \tau$, and fewer independent signatures remain available for constructing leave-one-client-out references. Nevertheless, CDRA continues to operate at $K = 320$, despite the substantial fragmentation of the original graph.

We additionally attempted the next doubled federation size, $K = 640$, but the Dirichlet partitioning procedure could not construct a valid allocation satisfying the required per-client support constraints. We therefore report $K = 320$ as the largest feasible configuration for this fixed graph and partitioning protocol, rather than as an architectural limit of CDRA. More generally, the practical scaling boundary depends jointly on the number of available nodes, the labeled training fraction, the Dirichlet concentration, the minimum-support threshold, and the resulting class-reference coverage. Computationally, signature collection and client prediction scale approximately linearly with K, while each uploaded structural signature remains independent of the local graph size and node-feature dimensionality.

H Efficiency Analysis

Table 7: Efficiency comparison on benchmark datasets for a single fixed seed. Training time is reported in seconds, peak GPU memory is reported in GB, and upload size is reported in MB. Best results within each metric group are shown in **bold**. Lower is better for all metrics. In the table header, G, O, and C denote GHOST, OASIS, and CDRA, respectively.

Dataset	Time ↓			GPU Mem. ↓			Upload ↓		
	G	O	C	G	O	C	G	O	C
Cora	38.17	26.93	**6.70**	0.21	0.28	**0.07**	15.80	44.57	**7.04**
CiteSeer	46.14	31.60	**8.04**	0.51	0.69	**0.16**	38.06	111.32	**18.12**
Chameleon	60.82	42.82	**19.62**	0.31	0.41	**0.10**	24.54	70.79	**11.38**
Actor	109.26	92.38	**13.44**	0.27	0.29	**0.09**	10.89	29.82	**4.58**
Roman-Empire	369.95	318.02	**18.52**	0.26	0.47	**0.09**	4.73	11.34	**1.56**
OGBN-Arxiv	12732.02	11453.43	**240.82**	16.53	28.61	**0.54**	3.15	6.49	**0.85**
Coauthor-CS	2414.36	2428.58	**40.96**	3.36	3.03	**0.85**	68.50	202.63	**33.31**

Table 7 provides the complete efficiency comparison of CDRA, GHOST, and OASIS across all seven datasets. We report runtime, peak GPU memory, and client-to-server upload size from a single fixed seed. These metrics capture complementary resource requirements: runtime measures computational cost, peak GPU memory reflects hardware demand, and upload size quantifies the communication required by the one-shot protocol.

CDRA achieves the lowest cost for every metric on every dataset. On the smaller graphs, its runtime advantage ranges from approximately $2.2\times$ to $20.0\times$ relative to GHOST and OASIS. The difference becomes substantially larger on the two largest computational settings. On OGBN-Arxiv, GHOST and OASIS require $52.9\times$ and $47.6\times$ more time than CDRA, respectively. On Coauthor-CS, the corresponding factors increase to approximately $58.9\times$ and $59.3\times$.

A similar pattern is observed for peak GPU memory. CDRA requires only 0.54 GB on OGBN-Arxiv, compared with 16.53 GB for GHOST and 28.61 GB for OASIS, corresponding to reductions of approximately $30.6\times$ and $53.0\times$. On the remaining datasets, CDRA also consistently uses less memory, with peak consumption remaining below 1 GB throughout the evaluation.

CDRA also produces the smallest upload payload on every dataset. Compared with CDRA, GHOST requires approximately $2.1\times$–$3.7\times$ more upload communication, while OASIS requires approximately $6.1\times$–$7.6\times$ more. This reduction follows from the form of the CDRA payload: clients transmit their trained models together with only eligible class-conditioned diffusion signatures and support counts. The structural component contains at most HC^2 floating-point values and C counts, independently of the number of nodes and node-feature dimension.

The efficiency gains arise from a fundamental difference in aggregation design. GHOST and OASIS rely on learned graph generators or synthesizers, server-side pseudo-graph construction, and additional knowledge-transfer optimization. CDRA instead computes its structural summaries through sparse local propagation and directly combines client predictions using class-specific reliability weights. The results therefore demonstrate that CDRA's predictive improvements do not require a more expensive server pipeline. Instead, the method simultaneously improves accuracy and substantially reduces computation, memory consumption, and communication overhead.

I Sensitivity to Diffusion Depth

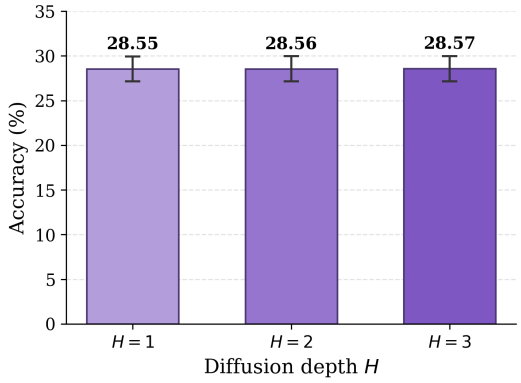


Figure 3: Sensitivity of CDRA to diffusion depth H on Actor. Bars show mean accuracy, and error bars indicate standard deviation over multiple random seeds.

Figure 3 evaluates the sensitivity of CDRA to the maximum diffusion depth H on the Actor dataset. Accuracy remains effectively unchanged as H increases from 1 to 3, varying only from 28.55% to 28.57%. This difference is negligible relative to the standard deviation of 1.4 percentage points. The results therefore indicate that CDRA is insensitive to the precise choice of H within the evaluated range. One-hop diffusion captures most of the useful class-mixing information on Actor, while additional depths provide largely redundant structural evidence. We retain $H = 3$ in the main experiments to incorporate multiple diffusion scales, with only linear growth in signature size and construction cost.

J Scope and Threat Model

CDRA addresses resource-efficient one-shot federated graph aggregation under statistical and structural heterogeneity. It estimates class-specific structural reliability without reconstructing client graphs, training auxiliary generators, or requiring iterative client-server communication.

We consider an honest-but-curious setting in which the server and participating clients follow the prescribed protocol. During the single communication round, the server receives the trained local model parameters θ_k, eligible class-conditioned diffusion signatures $\{\mathbf{S}_{k,c}\}_{c\in\mathcal{C}_k^\tau}$, and their support counts $\{n_{k,c}\}_{c\in\mathcal{C}_k^\tau}$. Raw node features, edges, node-level representations, and individual labels are retained locally and are not included in the payload.

The method does not claim formal differential privacy or protection against malicious clients, collusion, poisoning, or inference from uploaded models and aggregate summaries. Such adversarial security settings are outside the scope of this study. Our claims concern one-shot communication, compact structural summarization, class-wise reliability estimation, and computational, memory, and communication efficiency under non-IID graph partitions.

K Limitations

The principal methodological limitation of CDRA is its dependence on sufficient cross-client class support. A leave-one-client-out reference for class c requires at least two clients satisfying $n_{k,c} \geq \tau$. When a fixed graph is divided among a very large number of clients, or when label heterogeneity is particularly severe, some classes may lack an independent structural reference. CDRA handles such cases by assigning a neutral diffusion margin and retaining the client-size and class-support contributions. The threshold τ therefore controls a trade-off between class coverage and the reliability of locally estimated signatures.

The diffusion margin is most informative when classes exhibit distinguishable structural profiles. If several classes have nearly identical propagation behavior, or if the local graphs are extremely sparse, their signatures may provide limited separation. CDRA remains well defined in this regime and relies more heavily on its non-structural reliability components, but the additional benefit of diffusion-based evidence may be reduced.

CDRA further assumes a shared label space and compatible class probabilities across clients. Partially overlapping label spaces, heterogeneous prediction tasks, and open-set classes are not addressed by the current formulation. In addition, CDRA fuses the uploaded local models rather than producing a single global model. This avoids server-side retraining and graph reconstruction, but server storage and prediction costs increase with the number of participating clients.