BEYOND GROUP STATISTICS: LOCAL RELATIONSHIP PRIORS FOR PRIVATE GRAPH LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Private graph learning infers a propagation structure from noisy relational reports. Group statistics constrain how much connectivity is assigned to groups, but agreement with these statistics need not determine which local relations help a downstream task. We study this distinction through the Group–Local Prior (GLP). Given a fixed private report, GLP separates report-derived connection mass from a public-feature neighborhood preference, constructs a feasible mass-matched target, and pools the two priors before conditioning on randomized edge responses. An observation argument motivates the local preference, while normalized-propagation analysis explains why reconstruction fidelity need not predict the task effect. Experiments on six graphs, up to 89,250 nodes, evaluate this design under matched joint privacy budgets. GLP achieves the highest mean accuracy among evaluated private-report methods in 27 of 28 conditions on the original four-graph grid. On full Physics, it leads the evaluated private methods in both accuracy and Macro-F1 at both budgets and under both backbones; on Flickr, it improves three of four backbone–metric means over matched HPGR. Relationship controls and frozen-parameter confirmation connect these gains to local correspondence and task-relevant propagation. Together, these results support relationship placement as a task-oriented design principle for private graph learning.

1 INTRODUCTION

Graph neural networks (GNNs) use relationships to propagate information between nodes, but those relationships can be sensitive. In decentralized settings, an analyst may observe public node attributes while each participant locally randomizes their adjacency information. Randomized response (RR) offers link privacy, yet its false positive edges can overwhelm sparse graphs. Reconstruction methods therefore use additional information to recover useful structure: Blink combines noisy degrees with randomized edge reports, HPGR exploits homophily and group statistics, and SL-LPGL combines private cluster information with structure learning (Zhu et al., 2023; Xie et al., 2026; Zhang et al., 2024).

These approaches raise a design question: *given a fixed private report, how should the inferred connectivity be placed so that node features propagate usefully?* Group statistics describe connectivity to groups, but the resulting constraints need not determine the relevant local relationships. Figure 1 illustrates the distinction. Two graphs can have identical group counts and degrees while routing the same feature to different nodes. Edge responses supply further evidence, but their interpretation still depends on a structural prior. This suggests separating two roles: using the private report to infer an overall amount of connectivity, and using allowed local relationships to express a preference over its placement.

We develop the *Group–Local Prior* (GLP) around this separation.[1] Given the fixed private report, a group-based estimate first supplies a total probability mass M for the inferred connectivity. A neighborhood constructed from public features then supplies a local preference over where that mass may be placed. The core construction turns this preference into a *feasible, mass-matched target* T: a valid probability graph that keeps the report-inferred total mass, respects the support and capacity of

[1] Anonymous GLP code: https://anonymous.4open.science/r/sa24kjhkj123e.

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

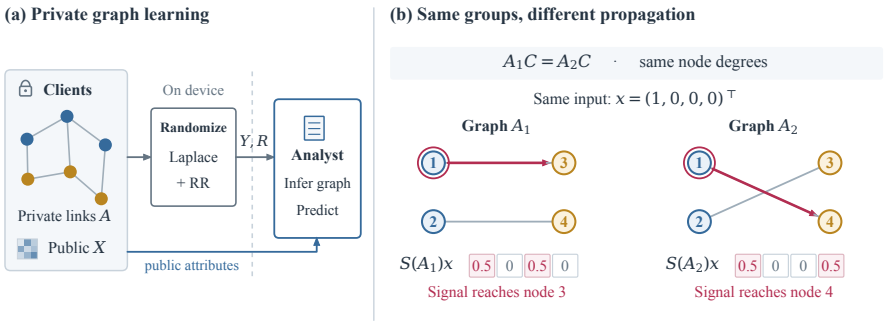


Figure 1: The setting and motivating distinction. (a) Attributes are public, relationships are private, and noisy group statistics Y and edge responses R are released once. (b) Blue and orange nodes define two groups. Both matchings have identical AC and degrees, but for $x = (1, 0, 0, 0)^\top$, propagation with unit self-loops sends its signal to different nodes. The ambiguity concerns the group channel; the full edge-response transcript supplies additional evidence.

the local neighborhood, and remains a well-defined object for the subsequent randomized-response conditioning. GLP pools T with the group prior *before* conditioning on the saved edge responses, so edge evidence reshapes the local placement rather than merely rescaling a fixed mixture. A capacity-aware branch also accommodates reports whose inferred mass exceeds what the local support can carry. The report is collected once and reused throughout.

The design is informed by goal-oriented inverse problems: estimating an object and estimating a task-relevant function of it can require different information (Stuart, 2010; Spantini et al., 2017; Adler et al., 2022). We make this distinction concrete for private graph learning. An observation-fiber calculation identifies changes that preserve group statistics but alter propagation. A report-to-task derivative explains how a prior correction interacts with edge evidence, degree normalization, and model sensitivity. Together, they motivate evaluating local relationship correspondence and task response alongside graph recovery.

Experiments cover six graphs and two backbones. On Cora, CiteSeer, Amazon Photo, and PubMed, GLP attains the highest mean accuracy among evaluated private-report methods in 27 of 28 backbone–dataset–budget conditions, with margins up to 8.27 pp over Blink Soft and 7.82 pp over SL-LPGL. Repeated full-graph experiments on Physics (34,493 nodes) and Flickr (89,250 nodes) extend the evidence beyond citation and product graphs. On Physics, GLP leads the evaluated private methods in both metrics under both backbones at both budgets. Flickr shows a more selective effect: GCN gains 1.13 pp in accuracy and 2.16 pp in Macro-F1 over matched HPGR, while GPR improves Macro-F1 but trails by 0.21 pp in accuracy. Relationship controls and frozen-parameter confirmation support local correspondence as a useful design choice, with gains conditioned on the graph, budget, and propagation model.

Our contributions are:

- A formulation of the distinction between group-level connectivity mass and task-relevant relationship placement, with explicit links to normalized propagation: an observation-fiber argument shows that group statistics leave placement undetermined, and a report-to-task derivative shows that reconstruction fidelity need not predict the task effect.
- A feasible, mass-matched local target with a capacity-aware construction, pooled with the group prior before randomized-response conditioning so that edge evidence reshapes the placement; the private report is reused without additional collection.
- A six-graph, two-backbone evaluation under matched joint privacy budgets, including repeated full-graph experiments up to 89,250 nodes, complemented by relationship and injection-position controls and frozen confirmation on new splits. These results identify both repeatable utility gains and conditions where graph recovery diverges from task utility.

2 Related work

Private graph reconstruction for GNNs. Blink uses Bayesian estimation of noisy links (Zhu et al., 2023); HPGR uses homophily-aware group information and degree correction (Xie et al., 2026); SL-LPGL combines private cluster information with iterative graph learning (Zhang et al., 2024). HoGS also uses feature-derived edge priors with randomized response (Xu et al., 2026). GLP builds on HPGR's report, group estimator, and edge-response update. Its additional design separates report-derived mass from local relationship preference and constructs a feasible target before conditioning on those responses. Public attributes, homophily, and Bayesian updating are established ingredients; our contribution concerns their specific use and the resulting connection to task-relevant propagation. Whole-system comparisons assess the resulting scheme, while same-report controls examine its added design. The common training protocols and implementation provenance are specified in Appendix B.

Learning graphs for a task. Graph structure learning recognizes that observed edges need not be the best computational graph for prediction (Franceschi et al., 2019; Chen et al., 2020). Goal-oriented inference and task-adapted reconstruction provide related views of what an observation must determine (Spantini et al., 2017; Adler et al., 2022). Probability pooling and variational updating provide tools for combining judgments and observations (Clemen & Winkler, 1999; Bissiri et al., 2016), while graph-filter stability describes sensitivity to operator perturbations (Gama et al., 2020). We use these established ideas to explain the decisions and measured behavior of a private reconstruction pipeline. The local preference construction and its empirical analysis are the focus, rather than a claim of new general pooling or perturbation identities.

3 Private reports and group–local priors

Setting and one-report privacy. Let $A \in \{0,1\}^{n\times n}$ be a symmetric, zero-diagonal adjacency matrix. Public side information consists of node identities, features X, and allowed training/validation labels and masks, fixed under the addition or removal of one undirected edge. A feature MLP supplies a fixed row-stochastic membership matrix $C \in \mathbb{R}^{n\times c}$; training nodes are anchored to their known classes. For a row-budget parameter e, let $\varepsilon_s = 0.1e$ and $\varepsilon_r = 0.9e$. Clients provide

$$Y = AC + Z, \quad Z_{ia} \overset{\text{ind}}{\sim} \text{Lap}(1/\varepsilon_s), \qquad \Pr(R_{ij} \neq A_{ij}) = (1+\exp\varepsilon_r)^{-1}, \quad i \neq j. \tag{1}$$

Binary randomized response (RR) follows Warner (1965); the channels and endpoint responses use independent randomness. The degree estimate is derived from $Y\mathbf{1}$, requiring no additional query.

Theorem 1 (Conditional edge privacy)**.** *One transcript (Y, R) is pure edge-DP with joint undirected-edge budget $B = 2(\varepsilon_s + \varepsilon_r) = 2e$. A single client's report is e-edge-LDP for a one-bit change in its adjacency row. Postprocessing this transcript with the fixed public information, including model training and validation selection, preserves the respective guarantees.*

The factor two accounts for both endpoints; all experimental axes use B. The guarantee protects edges conditional on the allowed information. Five benchmark repetitions estimate utility; releasing five independently collected transcripts of a private graph requires composition (Dwork & Roth, 2014). Appendix A.3 gives the proof and scope.

Overview. Figure 2 shows the full pipeline. Group statistics determine a group prior Q_g and its total mass M. Public features determine a local preference W. We combine M and W into a feasible target T, pool it with Q_g, and apply the saved RR evidence before forming a normalized propagation operator. Thus W is public, while T also depends on the report through M.

Report-derived mass and local preference. The group estimator computes $Q_g = \mathcal{G}(Y, C)$ before using R. We retain HPGR's degree-corrected group reconstruction and clip off-diagonal entries to $[0,1]$; Appendix B specifies the estimator and implementation. Separately, W is the nonnegative cosine top-10 neighborhood of raw public features, symmetrized by taking the union with maximum weights. Define

$$N = n(n-1), \quad M = \sum_{i\neq j}(Q_g)_{ij}, \quad \rho = M/N, \quad s = |\{(i,j) : W_{ij} > 0\}|, \quad K = us. \tag{2}$$

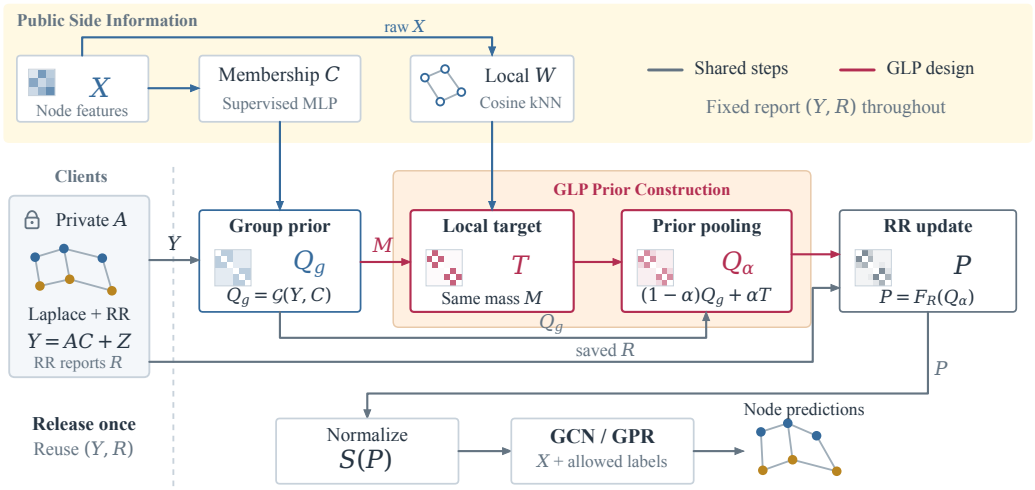


Figure 2: GLP reuses one report. Public features and allowed supervision define C; features also define a local neighborhood W. Group statistics yield Q_g and its mass M. The target T combines this report-derived mass with local preference. The pooled prior is conditioned on saved edge responses, normalized, and used by the classifier. Test labels enter only isolated evaluation.

Sums count ordered pairs, so M is inferred probability mass. The support capacity K uses the per-entry cap $u = 0.99$.

A feasible target at the inferred mass. We construct a symmetric target with zero diagonal and off-diagonal form

$$T_{ij} = \rho + a(V_{ij} - \bar{v}), \qquad \bar{v} = N^{-1}\sum_{i \neq j} V_{ij}. \tag{3}$$

When $0 < M < K$, choose $V = \min(bW, u)$ with $\sum V = M$ and set $a = 0.9$, giving $T = 0.9V + 0.1\rho$. When $K \leq M < N$ and the support is nonempty, use $V = u\mathbf{1}[W > 0]$ and $a = \min\{0.9, (1-\rho)/(\max V - \bar{v})\}$, with $a = 0.9$ if the denominator is zero. Empty support or $M \in \{0, N\}$ gives $T_{ij} = \rho$. Lemma 1 establishes $0 \leq T \leq 1$ and $\sum T = M$. A background carries mass beyond the capped support's capacity; the saturated branch retains support locations while dropping within-support weight ordering. This matches total prior mass, without imposing per-node degrees or group counts.

Pool, condition, and propagate. With $\alpha = 0.5$ frozen for the primary experiments, form

$$Q_\alpha = (1-\alpha)Q_g + \alpha T. \tag{4}$$

For $r_{ij} = R_{ij} + R_{ji}$, the two-endpoint likelihood ratio is $\lambda_{ij} = \exp\{2\varepsilon_r(r_{ij} - 1)\}$. Condition the pooled prior on these responses:

$$P_{ij} = F_{\lambda_{ij}}((Q_\alpha)_{ij}), \qquad F_\lambda(q) = \frac{\lambda q}{1 + (\lambda - 1)q}, \quad P_{ii} = 0. \tag{5}$$

Classifiers then use

$$S(P) = D^{-1/2}(P + I)D^{-1/2}, \qquad D = \mathrm{diag}((P + I)\mathbf{1}). \tag{6}$$

The response update can change the prior mass. Both Q_g and T depend on the report, so pooling treats them as structural judgments rather than independent evidence. Edgewise, equation (4) minimizes their weighted forward Bernoulli KL divergences to a common prior; equation (5) then minimizes reverse KL to that prior minus the observed log-likelihood term (Appendix A.4). These successive objectives explain the implemented ordering without implying a jointly optimal or calibrated graph posterior. The injection-position controls in Table 15 compare this placement with uniform smoothing and propagation-stage mixing. Collection, group reconstruction, and RR conditioning follow the shared pipeline; the feasible target and pooling define the added method.

4 FROM GROUP OBSERVATIONS TO TASK-RELEVANT PROPAGATION

What do group observations determine? The map $A \mapsto AC$ constrains adjacency through group-weighted sums. We first ask whether this information determines propagation in a local weighted-graph space.

Proposition 1 (Group agreement and propagation)**.** *Let C be row-stochastic and A symmetric with zero diagonal and off-diagonal entries strictly between zero and one. For any symmetric zero-diagonal H with $HC = 0$, all sufficiently small t satisfy*

$$(A + tH)C = AC, \qquad S(A + tH)X - S(A)X = tD_A^{-1/2} H D_A^{-1/2} X, \tag{7}$$

where $D_A = \operatorname{diag}((A + I)\mathbf{1})$. Thus the group observation locally determines this propagation exactly if and only if every such H also satisfies $D_A^{-1/2} H D_A^{-1/2} X = 0$.

Since $HC = 0$ implies $H\mathbf{1} = 0$, degrees are unchanged and equation (7) is exact. The design implication is that matching coarse connectivity alone need not settle where useful propagation occurs. GLP supplies a local preference to address this choice. The relationship-permutation experiment (Figure 3) and mass/change-norm control (Appendix E) test whether correspondence matters beyond adding weight. The proposition concerns the group channel in weighted space; full RR still distinguishes edge bits statistically, and GLP does not impose $(T - Q_g)C = 0$. Its correction therefore expresses a mass-matched preference, rather than exact null-space completion.

How a prior correction reaches the task. Fix a report and write $\Delta = T - Q_g$. Its task effect passes through both response conditioning and degree normalization.

Proposition 2 (Report-to-task response)**.** *Fix the model parameters θ. At a point where its loss $\ell_\theta(S)$ is differentiable,*

$$\frac{d\ell_\theta(S(P_\alpha))}{d\alpha} = \langle \nabla_S \ell_\theta, \mathcal{D}_{P_\alpha}[G_\alpha \odot \Delta] \rangle_F, \quad (G_\alpha)_{ij} = \frac{\lambda_{ij}}{[1 + (\lambda_{ij} - 1)(Q_\alpha)_{ij}]^2} \quad (i \neq j), \tag{8}$$

where the diagonal of $G_\alpha \odot \Delta$ is zero. For $d = (P + I)\mathbf{1}$ and $S = S(P)$,

$$\mathcal{D}_P[H] = D^{-1/2} H D^{-1/2} - \tfrac{1}{2} \operatorname{diag}(H\mathbf{1}/d)S - \tfrac{1}{2} S \operatorname{diag}(H\mathbf{1}/d). \tag{9}$$

A strictly negative derivative implies a loss decrease for a sufficiently small feasible positive step with this fixed model.

Equation (8) separates response weighting, normalization, and task sensitivity. Even a mass-preserving Δ can change individual degrees, so the last two terms of equation (9) matter. This motivates evaluating the final propagation effect alongside graph recovery: a reduction in an edge-recovery error need not align with a beneficial task direction. Figure 4 exhibits this separation empirically. The fixed-model diagnostic in Appendix D.3 further finds that including degree normalization reverses the local validation-loss direction relative to the direct edge term in 9 of 10 CiteSeer cases (Table 13). These results connect the derivative to measured behavior. Its scope is local loss at fixed parameters, without an accuracy guarantee after retraining or at the finite choice $\alpha = 0.5$. We use it for interpretation, not as an adaptive parameter selector.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

We evaluate six graphs with GPR and GCN backbones (Yang et al., 2016; Shchur et al., 2018; Chien et al., 2021; Kipf & Welling, 2017). Cora, CiteSeer, and Amazon Photo use joint budgets $B \in \{4, 6, 8, 12\}$; PubMed and Physics use $B \in \{4, 8\}$; Flickr uses $B = 4$. Each condition has five repetitions, except Flickr with three, on fixed partitions with approximately 10% training and 25% validation labels. We report means and sample SDs; repeated runs vary report and classifier randomness, not the data partition. GPR comparisons include Blink Soft/Hybrid and HPGR; GCN comparisons include SL-LPGL and HPGR on the original four graphs. The larger-graph comparison includes Blink under GPR on Physics and matched HPGR under both backbones on both graphs.

Table 1: Test accuracy (%, mean ± sample SD, five repetitions) on the original four graphs at $B = 4, 8$. Bold marks the highest evaluated private mean within each backbone, dataset, and budget. The non-private original graph is a separate reference reused across budgets.

Method	B	Cora	CiteSeer	Photo	PubMed
GPR					
GLP	4	$\mathbf{69.81 \pm 1.11}$	$\mathbf{70.17 \pm 0.64}$	$\mathbf{88.90 \pm 0.29}$	$\mathbf{83.28 \pm 0.27}$
GLP	8	$\mathbf{80.41 \pm 1.32}$	$\mathbf{69.93 \pm 1.27}$	$\mathbf{92.80 \pm 0.44}$	$\mathbf{84.31 \pm 0.18}$
HPGR	4	65.08 ± 0.65	68.27 ± 1.45	88.01 ± 0.71	82.47 ± 0.19
HPGR	8	78.14 ± 1.18	68.87 ± 0.32	92.49 ± 0.58	83.35 ± 0.33
Blink Soft	4	61.54 ± 0.74	63.22 ± 1.81	88.70 ± 0.31	82.61 ± 0.25
Blink Soft	8	78.55 ± 2.11	64.88 ± 1.52	92.54 ± 0.29	83.38 ± 0.35
Blink Hybrid	4	58.49 ± 1.84	56.88 ± 2.29	88.77 ± 0.63	81.34 ± 0.24
Blink Hybrid	8	74.98 ± 0.76	61.18 ± 1.79	91.62 ± 0.41	83.67 ± 0.16
Original graph	—	85.61 ± 0.71	73.03 ± 0.64	93.55 ± 0.41	86.69 ± 0.15
GCN					
GLP	4	$\mathbf{69.33 \pm 1.30}$	$\mathbf{70.18 \pm 0.92}$	$\mathbf{88.04 \pm 0.97}$	$\mathbf{84.61 \pm 0.26}$
GLP	8	$\mathbf{79.73 \pm 0.69}$	$\mathbf{70.16 \pm 0.85}$	$\mathbf{91.64 \pm 0.49}$	$\mathbf{85.58 \pm 0.24}$
HPGR	4	64.37 ± 1.00	68.22 ± 1.28	87.23 ± 0.87	83.43 ± 0.19
HPGR	8	76.81 ± 1.35	69.50 ± 0.79	91.35 ± 0.71	84.19 ± 0.35
SL-LPGL	4	68.93 ± 1.13	69.64 ± 1.05	84.92 ± 0.94	84.07 ± 0.23
SL-LPGL	8	71.76 ± 0.75	70.05 ± 0.77	84.57 ± 1.26	85.22 ± 0.24
Original graph	—	84.73 ± 0.85	72.14 ± 0.48	92.01 ± 0.15	87.42 ± 0.10

HPGR and GLP share reports and classifier seeds; different mechanisms retain their own queries at matched joint bounds. These are controlled comparisons, not original-paper table replications. Appendix B specifies the complete protocol, and Appendix C.1 documents large-graph coverage and resources.

5.2 Benchmark utility

Overall utility. Table 1 compares the original four graphs at their common budgets; Appendix C gives the wider grid and budget curves. GLP attains the highest mean accuracy among private-report methods in 27 of the 28 backbone–dataset–budget cells of the full grid; the single exception is Photo with GCN at $B = 6$, where HPGR leads by 0.38 pp. Margins vary with graph and budget. On Cora, GPR outperforms Blink Soft by 8.27/5.11/1.86 percentage points (pp) at $B = 4/6/8$ and HPGR by 4.73 pp at $B = 4$; on Photo, GLP–GCN exceeds full SL-LPGL by 3.12–7.82 pp across the four budgets. On CiteSeer the private methods are closely clustered: the margin over the closest private baseline is 0.53–2.12 pp at $B = 4, 6, 8$ and 0.24/0.37 pp at $B = 12$, with the smallest gap 0.11 pp (GCN at $B = 8$). Macro-F1 results are provided in Appendix C.

Full-graph results beyond 30,000 nodes. Table 2 adds Physics and Flickr without node subsampling or a sparse propagation approximation, retaining $\alpha = 0.5$ and $k = 10$. On Physics, GLP leads the evaluated private methods in all eight backbone–budget–metric means. Its GPR accuracy gains over Blink Soft are 0.79/0.47 pp at $B = 4/8$, and gains over Blink Hybrid are 1.40/1.35 pp. Relative to matched HPGR, both backbones improve accuracy by 0.45–0.85 pp, with all 20 paired changes positive. On Flickr, GCN improves accuracy and Macro-F1 by 1.13 and 2.16 pp; GPR improves Macro-F1 by 0.30 pp but loses 0.21 pp in accuracy. Thus the larger graphs support repeatable benefits while revealing a backbone-dependent boundary. Public-feature and original-graph references, full Macro-F1 results, and costs appear in Appendix C.1.

5.3 Relationship placement and task response

Where and what to inject. The injection-position controls use the original Cora study, where the three families select from the same weight grid by validation accuracy. Table 15 shows that, under the local family, the saved RR update after pooling outperforms direct operator mixing or replacing the target with a uniform graph: at 10% training labels, GLP reaches 79.44% while the

Table 2: Full-graph test accuracy (%, mean ± sample SD; five repetitions for Physics, three for Flickr). Bold marks the best evaluated private mean per column and backbone; dashes denote unmeasured conditions.

Method	Physics (34,493 nodes)		Flickr (89,250 nodes)
	$B = 4$	$B = 8$	$B = 4$
GPR			
GLP	$\mathbf{95.55 \pm 0.08}$	$\mathbf{95.96 \pm 0.15}$	45.56 ± 0.29
HPGR	94.72 ± 0.11	95.51 ± 0.08	$\mathbf{45.77 \pm 0.18}$
Blink Soft	94.76 ± 0.24	95.49 ± 0.11	—
Blink Hybrid	94.15 ± 0.21	94.61 ± 0.13	—
GCN			
GLP	$\mathbf{95.45 \pm 0.11}$	$\mathbf{95.89 \pm 0.08}$	$\mathbf{44.03 \pm 0.36}$
HPGR	94.60 ± 0.08	95.35 ± 0.07	42.89 ± 0.12

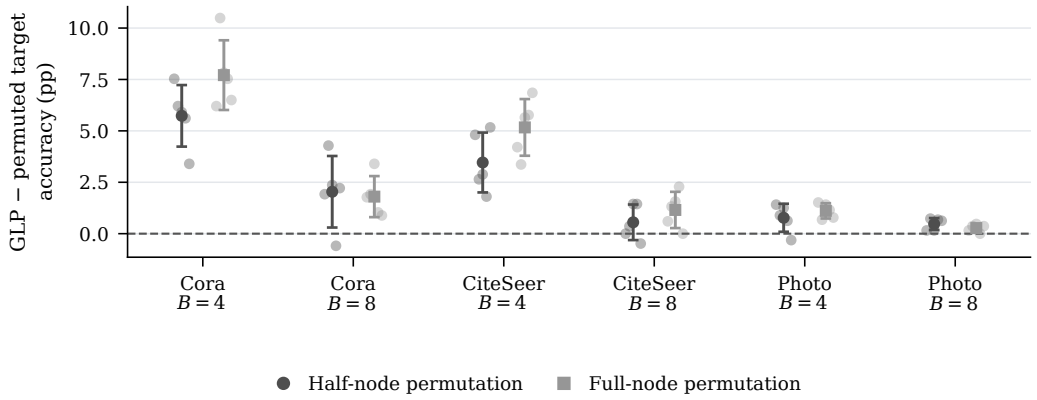


Figure 3: Paired GPR accuracy differences: GLP minus half- or full-node permutation controls, which preserve the target's off-diagonal entry multiset and mass. Faint points show individual runs; solid markers and error bars show five-run means and sample SDs. The dashed zero line denotes equal accuracy; positive values favor the intact local correspondence.

propagation and uniform families reach 78.35% and 78.20%, and the local family attains the highest selected mean at all three fractions. Appendix E.1 reports the measured selection grids, including a validation/test reversal in one propagation condition.

Relationship controls. Figure 3 compares GLP with targets whose node correspondence is partially or fully permuted while preserving the off-diagonal entry multiset and total mass. The intact target outperforms full permutation in all six graph–budget means and 29/30 paired GPR runs. The advantage shrinks as budget increases, from 7.71/5.17 pp on Cora/CiteSeer at $B = 4$ to 1.80/1.15 pp at $B = 8$; on Photo the $B = 8$ means are 0.27 pp apart. Appendix E adds matched-amplitude entry-permutation, group-density, and no-RR controls, including negative RR effects on CiteSeer. These controls measure the effect of the local construction under this protocol, not a universal causal share.

Recovery and task response. Figure 4 separates whole-graph Brier error and normalized membership-propagation error from GPR classification accuracy. It plots each error's per-run relative reduction of GLP over HPGR alongside the paired accuracy change. On the citation graphs the Brier and propagation-error reductions stay positive at every original budget, while the accuracy gain shrinks from 4.73 pp (Cora) and 1.90 pp (CiteSeer) at $B = 4$ to 0.24 and 0.34 pp at $B = 12$. On Photo the direction reverses for the membership-propagation error, which worsens in all 20 pairs. Appendix D.2 reports the full per-run values. The diagnostic supports a task-response interpretation of the local prior, not a claim that lower reconstruction error always improves downstream accuracy.

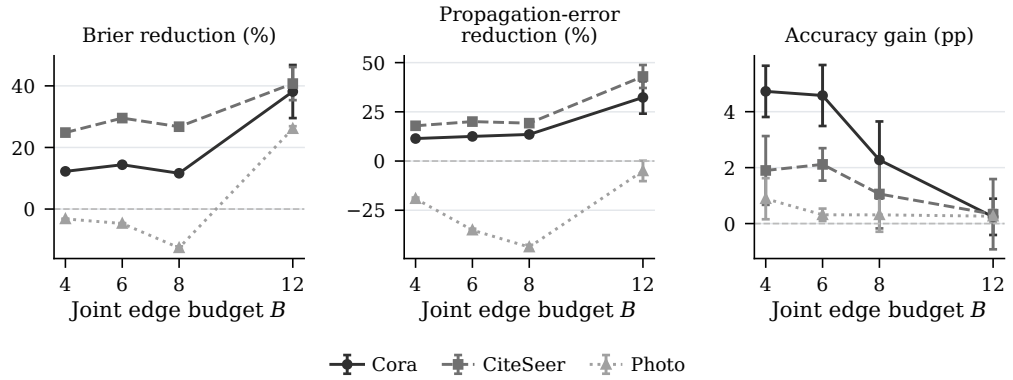


Figure 4: Structure recovery and task response relative to HPGR. Left: whole-graph Brier-error reduction. Middle: normalized membership-propagation-error reduction. Both use per-run relative reductions, $100(E_{\mathrm{HPGR}} - E_{\mathrm{GLP}})/E_{\mathrm{HPGR}}$ (Appendix D.2). Right: paired GPR accuracy gain in pp. Curves and error bars show five-run means and sample SDs. Dashed zero lines denote no change; positive values favor GLP.

5.4 Repeatability and scope

Frozen confirmation. With all method parameters frozen, new complete partitions and fresh reports on Cora/CiteSeer at $B = 4, 8$ give positive mean changes over matched HPGR in all eight graph–backbone–budget conditions for both metrics; 39/40 paired accuracy changes are positive (Appendix C.2). PubMed extends the complete comparison to 19,717 nodes without subsampling or sparse approximation: at both measured budgets, GLP attains the highest mean accuracy and Macro-F1 among private methods under both backbones.

Scope of the findings. Benefits depend on the graph, budget, and backbone. On CiteSeer the private-method margins are modest, falling to 0.24/0.37 pp at $B = 12$, and on Photo with GCN at $B = 6$, HPGR leads GLP by 0.38 pp in accuracy and 3.75 pp in Macro-F1. The public-feature reference remains competitive on the larger graphs, so the findings do not establish that private edges always add utility over public attributes. Full Flickr execution also requires substantial host memory: its first GLP reconstruction peaks at 223 GiB, and GPU training at about 30 GiB. Public features, allowed labels, fixed baseline configurations, and quadratic storage define the present scope. Complete results and costs appear in the appendices.

6 Conclusion

GLP separates report-derived connection mass from local relationship preference, constructs a feasible target, and pools it before conditioning on saved edge responses. Across six graphs and two backbones, the results support this placement-oriented design: the original four-graph grid, relationship controls, and frozen confirmation are complemented by repeated full-graph experiments on Physics and Flickr. These results extend the operating range to 89,250 nodes and show that the benefit can depend on the backbone and metric. The theory and diagnostics explain why reconstruction fidelity alone is insufficient to assess task utility. Public features, allowed labels, and quadratic storage remain explicit conditions; cross-prior transfer and memory-efficient implementations are directions for further work.

REPRODUCIBILITY STATEMENT

Appendix B describes the query, supervision, budget, implementation, and checkpoint-selection protocols; the appendices contain complete measured comparison grids and proofs, including the large-graph protocols and resource measurements in Appendix C.1. An anonymous implementation of GLP is available at https://anonymous.4open.science/r/sa24kjhkj123e.

ETHICS STATEMENT

We study conditional edge privacy on public benchmark graphs. Node identities, attributes, and permitted labels are treated as public side information; these assumptions do not establish privacy for people, attributes, labels, or whole neighborhoods. No new personal data were collected for this study.

AI USE STATEMENT

Generative AI tools were used for language polishing and grammar checking in this manuscript.

REFERENCES

Jonas Adler, Sebastian Lunz, Olivier Verdier, Carola-Bibiane Schönlieb, and Ozan Öktem. Task adapted reconstruction for inverse problems. *Inverse Problems*, 38(7):075006, 2022. doi: 10.1088/1361-6420/ac28ec. URL https://doi.org/10.1088/1361-6420/ac28ec.

Pier Giovanni Bissiri, Chris C. Holmes, and Stephen G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103–1130, 2016. doi: 10.1111/rssb.12158. URL https://arxiv.org/abs/1306.6430.

Yu Chen, Lingfei Wu, and Mohammed J. Zaki. Iterative deep graph learning for graph neural networks: Better and robust node embeddings. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/e05c7ba4e087beea9410929698dc41a6-Abstract.html.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized PageRank graph neural network. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2006.07988.

Robert T. Clemen and Robert L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999. doi: 10.1111/j.1539-6924.1999.tb00399.x. URL https://doi.org/10.1111/j.1539-6924.1999.tb00399.x.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. doi: 10.1561/0400000042. URL https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf.

Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning discrete structures for graph neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1972–1982. PMLR, 2019. URL https://proceedings.mlr.press/v97/franceschi19a.html.

Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability properties of graph neural networks. *IEEE Transactions on Signal Processing*, 68:5680–5695, 2020. doi: 10.1109/TSP.2020.3026980. URL https://arxiv.org/abs/1905.04497.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1609.02907.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110. URL https://econ-papers.upf.edu/papers/691.pdf.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018. URL https://arxiv.org/abs/1811.05868.

Alessio Spantini, Tiangang Cui, Karen Willcox, Luis Tenorio, and Youssef Marzouk. Goal-oriented optimal approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 39(5):S167–S196, 2017. doi: 10.1137/16M1082123. URL https://arxiv.org/abs/1607.01881.

Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061. URL https://authors.library.caltech.edu/records/fk9j9-5tq15.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965. doi: 10.1080/01621459.1965.10480775. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775.

Yule Xie, Jiaxin Ding, Pengyu Xue, Xin Ding, Haochen Han, Luoyi Fu, and Xinbin Wang. Leveraging homophily under local differential privacy for effective graph neural networks. In *Machine Learning and Knowledge Discovery in Databases. Research Track (ECML PKDD 2025)*, volume 16017 of *Lecture Notes in Computer Science*, pp. 380–396. Springer, 2026. doi: 10.1007/978-3-032-06096-9_22. URL https://link.springer.com/chapter/10.1007/978-3-032-06096-9_22.

Wen Xu, Zhetao Li, Yong Xiao, Pengpeng Qiao, Mianxiong Dong, and Kaoru Ota. HoGS: Homophily-oriented graph synthesis for local differentially private GNN training. *arXiv preprint arXiv:2602.08762*, 2026. URL https://arxiv.org/abs/2602.08762v1.

Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 40–48. PMLR, 2016. URL https://proceedings.mlr.press/v48/yanga16.html.

Guanhong Zhang, Xiang Cheng, Jiaan Pan, Zihan Lin, and Zhaofeng He. Locally differentially private graph learning on decentralized social graph. *Knowledge-Based Systems*, 304:112488, 2024. doi: 10.1016/j.knosys.2024.112488. URL https://www.sciencedirect.com/science/article/pii/S0950705124011225.

Xiaochen Zhu, Vincent Y. F. Tan, and Xiaokui Xiao. Blink: Link local differential privacy in graph neural networks via Bayesian estimation. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pp. 2651–2664. ACM, 2023. doi: 10.1145/3576915.3623165. URL https://arxiv.org/abs/2309.03190.

A Proofs and Additional Analysis

Throughout this appendix, graph matrices are symmetric with zero diagonal before self-loops are added. A probability graph has entries in $[0, 1]$. We use the ordered off-diagonal index set $\Omega = \{(i, j) : i \neq j\}$ and $N = |\Omega| = n(n-1)$, so a total mass counts each undirected edge twice. The results concern the mathematical mechanism in exact arithmetic; the finite-precision implementation is evaluated separately. All task derivatives hold for a fixed report and fixed model parameters.

A.1 Capacity-aware local preference

Here we give the complete definition behind Eq. (3), including boundary cases. Let Q_g be a probability graph and let W be finite, nonnegative, symmetric, and zero-diagonal. Define

$$M = \sum_{\Omega} (Q_g)_{ij}, \qquad \rho = M/N, \qquad s = |\{(i, j) : W_{ij} > 0\}|, \qquad K = us,$$

where $n \geq 2$, $u \in (0, 1)$, and $\eta \in (0, 1)$. Our fixed settings are $u = .99$ and $\eta = .9$. If $M \in \{0, N\}$ or $s = 0$, set $T_{ij} = \rho$ on Ω. Otherwise, construct T using one of the following branches:

$$0 < M < K: \quad V_{ij} = \min\{bW_{ij}, u\}, \quad \sum_{\Omega} V_{ij} = M, \quad T_{ij} = \eta V_{ij} + (1 - \eta)\rho; \tag{10}$$

$$M \geq K: \quad V_{ij} = u\,\mathbf{1}\{W_{ij} > 0\}, \quad \bar{v} = K/N, \quad h = \max_{\Omega} V_{ij} - \bar{v},$$

$$a = \begin{cases} \min\{\eta, (1 - \rho)/h\}, & h > 0, \\ \eta, & h = 0, \end{cases} \qquad T_{ij} = \rho + a(V_{ij} - \bar{v}). \tag{11}$$

All entries specified in these equations are off-diagonal; finally set $T_{ii} = 0$. In the unsaturated branch the same centered expression holds with $a = \eta$ and $\bar{v} = M/N$.

Lemma 1 (Feasibility and mass preservation)**.** *The target above is well defined, symmetric, and entrywise in* $[0, 1]$*, with* $\sum_{\Omega} T_{ij} = M$*. Consequently, Eq.* (4) *defines a probability graph with mass* M *for every* $\alpha \in [0, 1]$*.*

Proof. The boundary branch is immediate. For $0 < M < K$, the function $b \mapsto \sum_{\Omega} \min\{bW_{ij}, u\}$ is continuous and nondecreasing, starts at zero, and reaches K at a finite b: the finite nonempty positive support has a positive minimum weight. Hence a suitable b exists. Since $V \in [0, u]$ and $\sum_{\Omega} V = M$, Eq. (10) has total mass M and entries bounded above by $\eta u + (1 - \eta)\rho \leq 1$.

In the saturated branch, $0 \leq \bar{v} \leq \rho \leq 1$ and $0 \leq a \leq \eta < 1$. Therefore

$$T_{ij} \geq \rho - a\bar{v} \geq (1 - a)\rho \geq 0.$$

For $h > 0$, the definition of a gives $T_{ij} \leq \rho + ah \leq 1$. If $h = 0$, all off-diagonal entries of V equal their average, so $T_{ij} = \rho$. The centered term has zero sum, proving $\sum_{\Omega} T = M$. Symmetry follows from W. Convex pooling with Q_g preserves all these properties. □

Empty support and extreme masses thus require no additional private query. Full support in the saturated branch gives $h = 0$ and a uniform target. At $M = K$, the two nonboundary branches agree by continuity. In the saturation regime, V retains support locations but loses within-support weight ordering; the background carries mass that the capped support cannot carry alone. The claim concerns prior mass: the RR update need not preserve it. The finite bisection used to obtain b satisfies numerical tolerance rather than a bitwise mass identity. Also, while W is public, T depends on the report through M and is not statistically independent of Q_g.

A.2 Group observations and task-relevant ambiguity

Proof of Proposition 1. Let C be row-stochastic and let H be symmetric and zero-diagonal with $HC = 0$. Since $C\mathbf{1} = \mathbf{1}$, we also have $H\mathbf{1} = 0$. Thus $A + tH$ and A have the same group observation and the same degree vector. With $D_A = \mathrm{diag}((A + I)\mathbf{1})$, direct subtraction gives

$$(A + tH)C = AC, \qquad S(A + tH)X - S(A)X = tD_A^{-1/2} H D_A^{-1/2} X.$$

For an interior weighted graph A, both signs of every sufficiently small finite perturbation are feasible. Define the linear maps on symmetric zero-diagonal matrices by $\mathcal{O}_C(H) = HC$ and $\mathcal{J}_{A,X}(H) = D_A^{-1/2} H D_A^{-1/2} X$. The displayed identity shows that propagation is constant along every locally feasible observation fiber if and only if $\ker \mathcal{O}_C \subseteq \ker \mathcal{J}_{A,X}$. □

This ambiguity also occurs for binary graphs. For four nodes, choose $C = \mathbf{1}$ and $X = (1, 0, 0, 0)^\top$. Let A_0 contain edges $\{1, 2\}, \{3, 4\}$ and A_1 contain edges $\{1, 4\}, \{2, 3\}$. Both satisfy $A_i C = \mathbf{1}$, but

$$S(A_0)X = (1/2, 1/2, 0, 0)^\top, \qquad S(A_1)X = (1/2, 0, 0, 1/2)^\top.$$

The vector space of symmetric zero-diagonal real matrices has dimension $n(n-1)/2$, so its group-observation kernel has dimension at least $\max\{0, n(n-1)/2 - nc\}$. This dimension statement is about weighted matrices; it does not imply binary-graph ambiguity for every soft C.

Equal AC implies equal distributions of the noisy group channel. It does not imply equal distributions of the complete report, because positive-budget RR distinguishes different edge bits statistically. Nor does GLP impose $(T - Q_g)C = 0$. Proposition 1 motivates additional relational preferences; it does not identify the measured gains as exact null-space recovery.

A.3 Conditional edge privacy

Proof of Theorem 1. Condition on fixed node identities, features, allowed supervision and masks, and a row-stochastic C obtained without private-edge access. Toggling one undirected edge $\{i, j\}$ changes only rows i and j of AC, by $\pm C_j$ and $\pm C_i$. Their combined ℓ_1 norm is two. Consequently, independent Laplace noise of scale $1/\varepsilon_s$ gives a density ratio at most $\exp(2\varepsilon_s)$ for the complete group report Y.

Only the two ordered input bits A_{ij} and A_{ji} change in the RR channel. Each binary response with flip probability $(1 + \exp \varepsilon_r)^{-1}$ has likelihood ratio at most $\exp \varepsilon_r$. Independent responses therefore give a ratio at most $\exp(2\varepsilon_r)$. Independence of the two channels bounds the joint likelihood ratio by $\exp(2(\varepsilon_s + \varepsilon_r))$. The same calculation for a single adjacency row contains one changed group vector and one changed RR bit, giving local budget $\varepsilon_s + \varepsilon_r$. With $\varepsilon_s = .1e$ and $\varepsilon_r = .9e$, the complete undirected transcript has budget $B = 2e$.

The degree estimate $Y\mathbf{1}$, group reconstruction, local target, RR-conditioned graph, normalization, and any training or selection that uses only this report and the fixed allowed information are postprocessing. They add no edge-privacy cost. □

The guarantee does not protect node presence, attributes, labels, or an entire adjacency row under arbitrary changes. It is per report: joint release of independently recollected reports requires composition. Noise-generation seeds saved in the private experimental archive are not part of a publicly releasable DP transcript. The proof assumes ideal noise sampling and does not certify a particular PRNG or floating-point implementation for deployment on sensitive data.

For completeness, the group inverse exists whenever C has full column rank and its clipped degree weights satisfy $d_i \geq 1$: for nonzero z,

$$z^\top C^\top \operatorname{diag}(d) C z = \sum_i d_i (C_i z)^2 > 0.$$

One one-hot training anchor per class suffices for full column rank. This is not a floating-point conditioning guarantee. A failure symbol depending only on allowed inputs remains postprocessing; no untested inverse fallback is needed for the privacy argument.

A.4 Sequential variational interpretation

The implemented arithmetic pool and RR update admit two successive edgewise objectives. We state the characterization here to make the ordering precise.

Proposition 3 (Pooling followed by evidence). *For $g, t \in [0,1]$, $\alpha \in [0,1]$, and $\lambda > 0$, the updates in equations (4)–(5) are the successive minimizers*

$$q_\alpha = \arg\min_{q\in[0,1]} (1-\alpha)\,\mathrm{KL}(\mathrm{Ber}(g)\|\,\mathrm{Ber}(q)) + \alpha\,\mathrm{KL}(\mathrm{Ber}(t)\|\,\mathrm{Ber}(q)), \tag{12}$$

$$p_\alpha = \arg\min_{p\in[0,1]} \mathrm{KL}(\mathrm{Ber}(p)\|\,\mathrm{Ber}(q_\alpha)) - p\log\lambda. \tag{13}$$

The feasible targets in equation (3) ensure that matrix pooling preserves symmetry, bounds, and total prior mass.

Proof of Proposition 3. For one edge write $g = (Q_g)_{ij}$, $t = T_{ij}$, and $q_\alpha = (1-\alpha)g + \alpha t$. After removing terms independent of q, the forward-KL pooling objective becomes

$$-q_\alpha \log q - (1-q_\alpha)\log(1-q).$$

Its unique minimizer is $q = q_\alpha$. At boundary probabilities, use the extended-real KL convention; if $\alpha = 0$ or 1, omit the zero-weight term.

Next, for $0 < q_\alpha < 1$, differentiating $\mathrm{KL}(\mathrm{Ber}(p)\|\,\mathrm{Ber}(q_\alpha)) - p\log\lambda$ with respect to p gives

$$\mathrm{logit}(p) - \mathrm{logit}(q_\alpha) - \log\lambda.$$

Its second derivative is $1/[p(1-p)] > 0$, and the stationary point is $p = \lambda q_\alpha/(1 - q_\alpha + \lambda q_\alpha)$, as in Eq. (5). For $q_\alpha = 0$ or 1, finite KL forces $p = q_\alpha$, again agreeing with this expression. For two endpoint responses with sum $r = R_{ij} + R_{ji}$, their independent edge/nonedge likelihood ratio is $\lambda = \exp(2\varepsilon_r(r-1))$. □

The first KL direction yields arithmetic pooling; reversing that direction generally yields a log-odds pool. These are successive variational problems, not a demonstrated joint optimum. The second is an edgewise update under a working prior, not a calibrated posterior over all graphs. No independence of T and Q_g or optimality of the fixed $\alpha = .5$ is implied.

A.5 From prior corrections to task responses

Proof of Proposition 2. For fixed $\lambda > 0$,

$$F'_\lambda(q) = \frac{\lambda}{[1+(\lambda-1)q]^2}.$$

Thus, along $Q_\alpha = Q_g + \alpha\Delta$ with $\Delta = T - Q_g$, the off-diagonal posterior derivative is $H = \dot{P}_\alpha = G_\alpha \odot \Delta$, with diagonal zero. Let $d = (P+I)\mathbf{1}$, $D = \mathrm{diag}(d)$, and $h = H\mathbf{1}$. Differentiating $d_i^{-1/2}$ gives $-h_i/(2d_i^{3/2})$. Applying the product rule to Eq. (6) therefore yields

$$\mathrm{D}S_P[H] = D^{-1/2}HD^{-1/2} - \tfrac{1}{2}\,\mathrm{diag}(h/d)S - \tfrac{1}{2}S\,\mathrm{diag}(h/d).$$

Unit self-loops ensure $d_i \geq 1$. Applying the chain rule to a differentiable fixed-model loss gives

$$\frac{d\ell_\theta(S_\alpha)}{d\alpha} = \langle \nabla_S \ell_\theta, \mathrm{D}S_{P_\alpha}[G_\alpha \odot \Delta]\rangle_F.$$

A strictly negative derivative implies decrease for sufficiently small positive feasible steps. If $f(\alpha) = \ell_\theta(S_\alpha)$ has L-Lipschitz derivative on $[0,a]$, integrating $f'(t) \leq f'(0) + Lt$ gives

$$f(a) - f(0) \leq af'(0) + \tfrac{1}{2}La^2.$$

□

An equivalent edgewise task representation follows by setting $J = \mathrm{sym}(\nabla_S \ell_\theta)$, $E = D^{-1/2}$, and $v_i = \sum_j J_{ij}S_{ij}/d_i$:

$$K_\theta = EJE - \tfrac{1}{2}(v\mathbf{1}^\top + \mathbf{1}v^\top), \qquad f'(\alpha) = \langle K_\theta, G_\alpha \odot \Delta\rangle_F.$$

Indeed, the degree terms combine to $-v^\top H\mathbf{1}$, equal to the inner product with the second term of K_θ for symmetric H. The Frobenius inner product already includes both positions of each undirected edge.

Global mass preservation does not imply $H\mathbf{1} = 0$, so the degree terms cannot generally be omitted. ReLU models require differentiable regions or an appropriate piecewise argument; Lipschitz differentiability must not be assumed across arbitrary kinks. The result does not guarantee improvement at $\alpha = .5$, in test accuracy, or after separately retraining parameters.

For two prior values, direct subtraction also gives

$$F_\lambda(q_1) - F_\lambda(q_0) = \frac{\lambda(q_1 - q_0)}{[1 + (\lambda - 1)q_0][1 + (\lambda - 1)q_1]}.$$

RR preserves the sign of a single-edge correction but changes its magnitude. For fixed internal priors, the multiplier tends to zero as $\lambda \to 0$ or ∞, and equals one at $\lambda = 1$. This is not a monotone budget law: the report, group estimate, target, and response categories all change across budgets. Boundary priors remain locked at zero or one.

A.6 Risk and margin interpretations

Lemma 2 (Complementarity in a linear risk proxy)**.** *For a fixed linear map L, a fixed target Q_*, and square-integrable errors $e_g = L(Q_g - Q_*)$, $e_t = L(T - Q_*)$, define $R_g = \mathbb{E}\|e_g\|^2$, $R_t = \mathbb{E}\|e_t\|^2$, and $C_{gt} = \mathbb{E}\langle e_g, e_t\rangle$. Then*

$$R(\alpha) = (1 - \alpha)^2 R_g + \alpha^2 R_t + 2\alpha(1 - \alpha)C_{gt}.$$

If $D = R_g + R_t - 2C_{gt} > 0$, its minimizer on $[0, 1]$ is $\mathrm{clip}_{[0,1]}((R_g - C_{gt})/D)$. If $C_{gt} < R_g$, sufficiently small positive mixing improves this proxy risk; if $C_{gt} < \min\{R_g, R_t\}$, the interior optimum strictly improves both endpoints.

Proof. Expand the squared norm of $(1 - \alpha)e_g + \alpha e_t$, take expectations, and minimize the quadratic. Its derivative at zero is $2(C_{gt} - R_g)$, and $D = \mathbb{E}\|e_t - e_g\|^2$. When $D = 0$, the two errors agree almost surely and all mixing coefficients have the same risk. □

This is a shrinkage-style risk calculation (Ledoit & Wolf, 2004) that allows biased, correlated errors. In particular, C_{gt} is an uncentered cross-error moment and cannot be removed for our report-dependent target. The optimal coefficient is an oracle expression, not an implemented estimator or an accuracy guarantee for the nonlinear, retrained pipeline.

Lemma 3 (Fixed-GPR perturbation bound)**.** *Let fixed logits be $Z(S) = \sum_{k=0}^{K} \gamma_k S^k U$, with fixed U and coefficients γ_k. For symmetric normalized nonnegative graph operators S, S',*

$$\|Z(S') - Z(S)\|_F \le \left(\sum_{k=1}^{K} k|\gamma_k|\right) \|S' - S\|_2 \|U\|_F.$$

Proof. Such operators are similar to row-stochastic matrices and symmetric, so their spectral norms are at most one. The identity $S'^k - S^k = \sum_{j=0}^{k-1} S'^j (S' - S) S^{k-1-j}$, followed by submultiplicativity and the triangle inequality, gives the bound without assuming that S and S' commute. □

For a node with positive true-class margin $m_i = Z_{iy_i} - \max_{k \neq y_i} Z_{ik}$, let $\delta_i = \|Z_i(S') - Z_i(S)\|_\infty$. If $m_i > 2\delta_i$, its correct prediction is preserved, since any class-pair margin decreases by at most $2\delta_i$. Correcting a wrong prediction additionally requires a favorable direction. Public-membership entropy is not the same quantity as a GNN classification margin, and this fixed-parameter bound does not cover separately learned feature networks or GPR coefficients. Together, these conditions motivate examining correction opportunities and disruption risks without asserting that structural fidelity, uncertainty, or stability alone determines utility.

B Implementation and evaluation protocol

B.1 Datasets and supervision

Table 3 describes the processed graphs. Edges are unique undirected non-self links; duplicate directed entries and self-loops are removed before constructing the symmetric adjacency. This matters

for CiteSeer: the processed graph contains 9,104 directed positions, rather than treating a raw-file edge count as the number used in these experiments. Features are the dataset attributes. Public W is constructed from raw features, not the hidden representation of the membership MLP. The first

Table 3: Processed graph sizes and original split counts. Nodes outside training, validation, and test remain available as unlabeled transductive nodes.

Dataset	Nodes	Edges	Features	Classes	Train/val/test	Unlabeled
Cora	2,708	5,278	1,433	7	270/677/677	1,084
CiteSeer	3,327	4,552	3,703	6	332/832/832	1,331
Photo	7,650	119,081	745	8	765/1,912/1,912	3,061
PubMed	19,717	44,324	500	3	1,971/4,929/4,929	7,888

three datasets use $B = 4, 6, 8, 12$; PubMed uses $B = 4, 8$. PubMed is processed in full, without node/feature subsampling or a sparse similarity approximation. Its sequential stratified partition uses initial seed 2026091401 and the same 10/25/25/40% target proportions, rather than the standard public Planetoid split. The stored PubMed adjacency has 88,648 directed positions, corresponding to the 44,324 undirected edges in the table.

The confirmatory partitions are generated sequentially with stratification: 10% training, then 25% validation, then 25% test, with the remaining approximately 40% unlabeled. The initial split seeds are 2026091311 (Cora) and 2026091321 (CiteSeer); the second and third stages increment these by one and two. The resulting counts are 270/677/677/1084 and 332/831/832/1332. The split constructor uses benchmark labels for stratification, after which training/validation supervision is stored separately and test labels are isolated from reconstruction, training, and model selection. These are new partitions of the same graphs, with node overlap across old and new splits.

A feature-only MLP is trained once per dataset/split for the HPGR membership matrix. Its saved log probabilities are temperature-scaled with $\tau = 0.75$; nontraining nodes use $C_i = \mathrm{softmax}(\log p_i/\tau)$. Training nodes alone receive one-hot anchors. Validation labels select the MLP checkpoint but are not used as one-hot anchors. Thus the 10% training-label count is not the total amount of allowed supervision. A shared C is then fixed before the repeated private reports are generated. The five runs do not retrain five independent membership models. PubMed fixes membership seed 26001 and classifier seeds 26010–26014; the separate SL cluster MLP uses seed 27001. The same classifier seed index across mechanisms does not make their report noise identical.

B.2 The matched group estimator

Set $d_i = \mathrm{clip}((Y\mathbf{1})_i, 1, n)$, $g_a = \sum_i C_{ia}$, and $H = C^\top Y$. Following the frozen HPGR estimator, let $v_a = 2g_a/\varepsilon_s^2$ and symmetrize the group matrix as

$$\widehat{H}_{aa} = H_{aa}, \qquad \widehat{H}_{ab} = \widehat{H}_{ba} = \frac{H_{ab}/v_a + H_{ba}/v_b}{1/v_a + 1/v_b} \quad (a < b). \tag{14}$$

Then compute

$$U = C^\top \mathrm{diag}(d) C, \quad \Theta = U^{-1}\widehat{H}U^{-1}, \quad Q_g = \mathrm{offdiag}\big(\mathrm{clip}\left(\mathrm{diag}(d)C\Theta C^\top \mathrm{diag}(d), 0, 1\right)\big). \tag{15}$$

This specifies the estimator used here without asserting that its soft-membership symmetrization is an exact maximum-likelihood estimate. With at least one one-hot anchor in every class and $d_i \geq 1$, U is positive definite. The implementation uses a direct inverse, not an added ridge or pseudoinverse; inputs violating its requirements can return a failure symbol as part of the postprocessing definition. The target's unsaturated scale is computed with 50 bisection steps. In the archived extension, its maximum relative mass discrepancy was 1.70×10^{-7}.

B.3 Backbones and frozen parameters

Both backbones use hidden dimension 16 and 500 training epochs. Every 20 epochs, validation accuracy is computed in evaluation mode; the first best checkpoint is retained. Test metrics are calculated only after configuration selection and checkpoint saving.

GPR. The feature network is a two-layer MLP with ReLU and dropout 0.4 at its input and hidden layer. Its logits are propagated by $\sum_{k=0}^{10} \gamma_k S^k$, with trainable coefficients initialized to $\gamma_k = 0.1(0.9)^k$ for $k < 10$ and $\gamma_{10} = (0.9)^{10}$. Adam uses learning rate 0.001 and weight decay 5×10^{-4} on the feature layers, with zero weight decay on the propagation coefficients. The frozen graph-training objective uses summed NLL; this differs from the mean-NLL convention of some original GPR training scripts and is shared across the GPR comparisons. The membership MLP uses mean NLL.

GCN. The fixed-graph network is $S\,\mathrm{Dropout}(\mathrm{ReLU}(SXW_1))W_2$, followed by log-softmax, with bias-free linear layers and hidden dropout 0.5. Adam uses learning rate 0.01, weight decay 5×10^{-4}, and mean NLL. SL-LPGL retains its additional learned-graph and feature-enhancement modules and its composite loss, rather than being reduced to the fixed-graph network.

Method parameters. All primary GLP results use $\alpha = 0.5$, $u = 0.99$, $a \leq 0.9$, public cosine $k = 10$, temperature 0.75, and statistical/RR allocation 0.1/0.9. The capacity construction was frozen before the new budget extension and complete-split confirmation. The original Cora weight grid and selection rule are reported in Appendix E.1; no later dataset or new complete partition reselects the GLP mixing weight. PubMed retains these parameters and the frozen SL/Blink configurations. Physics and Flickr also retain the GLP parameters; their measured baseline coverage and randomization settings are listed in Appendix C.1. The original-graph reference is a non-private comparator under the same backbone protocol, not a theoretical upper bound: a learned or attribute-derived graph can be better suited to the labels.

B.4 Baseline provenance and comparison scope

HPGR. The group reconstruction and two-endpoint response logic are based on the author repository `https://github.com/xyl-alter/HPGR`, commit `e644d1f`. The matched baseline adds $[0, 1]$ clipping and zeroing the prior diagonal before the response update and uses our common membership/label and backbone protocol. The upstream prior-return path only lower-clips. Accordingly, the table label HPGR denotes this matched implementation. Earlier Cora author-protocol runs gave six means within approximately 0.48 pp of the repository's saved reference means; that anchor does not make the present changed-protocol tables reproductions of the paper's numerical tables.

Blink. We use the author query/reconstruction modules from `https://github.com/zhxchd/Blink_GNN`, commit `e55cb7b`. Soft and Hybrid reuse their own saved reports and are evaluated with our common GPR classifier. The degree-statistic/RR budget ratio is fixed at 0.1/0.9. We do not substitute HPGR's group query for Blink's degree query. This is a transfer of the author reconstruction modules, not a rerun of the original full training system or its hyperparameter search.

SL-LPGL. The complete pipeline is reconstructed in this project from Zhang et al. (2024), including HAGEI, EADEGSL, and joint GCN optimization. The IDGL author implementation (Chen et al., 2020) is an upstream reference for graph-learning operations, not SL-LPGL author code. The private report contains hard-cluster degrees plus independent Laplace noise of scale $2/B$, yielding the same undirected-edge bound. Its cluster MLP uses hidden dimension 16 and dropout 0.5; hard clusters and hidden embeddings come from its saved checkpoint without test-label access.

The fixed SL configuration uses $k = 10$, $\rho = 0.1$, $\psi = 0.9$, $\lambda = 0.8$, $\eta = 0.1$, $\xi = 4 \times 10^{-5}$, and regularization parameters $(0.1, 0.05, 0.1)$, in the original method's notation. Only $\delta \in \{0.1, 0.3, 0.5\}$ was initially compared by mean validation accuracy on Cora; $\delta = 0.5$ was then frozen for transfer. There are at most ten graph-learning iterations per forward pass. The training loss is the first-iteration loss plus the mean of later-iteration losses, or just the first when there is one iteration.

Implementation conventions include stable node-index tie breaking, rounded clipped cluster counts for HAGEI, symmetrization by $(A + A^\top)/2$, nonnegative row-normalized learned cosine graphs, and a stopping condition based on successive graph change relative to the first graph. Empty training neighborhoods fall back to the node's own report; test-time neighborhoods use top-k candidates. These declared conventions resolve underspecified cases. SL's fixed-configuration transfer and additional trainable modules limit attributing a whole-system gap to one prior component.

The comparisons therefore answer a controlled question about these implementations and budgets. They do not establish universally optimized baseline performance. The author-module origins, paper-based reconstruction, and current training protocols remain distinct, even when a checkpoint exactly reproduces its saved output.

B.5 Reports, repetitions, and computational record

Each report is persisted before postprocessing with its query definition, membership matrix, split/input identifiers, budget allocation, random seeds, code version, and content hashes. HPGR and GLP read the same report; Blink variants share the corresponding Blink report. Different query types, budgets, or memberships produce distinct versions. The same repetition number across mechanisms is not a claim that their noises are identical.

The three-graph consolidated extension contains 680 main/comparison records: 470 new classifier trainings and 210 reused records. Its 130 new reports comprise 50 HPGR, 40 Blink, and 40 SL reports; three additional MLP trainings prepare memberships. PubMed contributes 90 distinct classifier models: 60 initial HPGR/GLP/reference models and 30 additional Blink/SL models. The external-baseline stage reuses the earlier 60 records without retraining. Across both PubMed stages, the 30 private reports comprise 10 HPGR, 10 Blink, and 10 SL reports, prepared with two auxiliary membership/cluster MLPs. Earlier stages contain component and directional controls described separately below. Counts of checkpoints or metrics are not counts of independent graphs. Primary-grid SDs are sample SDs over five runs; earlier subset controls explicitly identify their 15-observation aggregation. They are not standard errors or confidence intervals; we do not infer population-level significance from repeated nodes or graphs.

The original four-graph experiments ran on an RTX 5090 with 32 GB VRAM and 64 GB host RAM, Python 3.11.14, PyTorch 2.9.1+cu130, and CUDA 13.0 under Linux. The three-graph extension's recorded classifier-training time is 1,222.91 seconds and its peak CUDA allocated memory is approximately 5.35 GiB. These figures exclude collection, reconstruction, I/O, transfer, and auditing.

The initial PubMed phase prepares one HPGR membership model, the public-feature operator W, ten reports, and twenty private propagation graphs, and trains and audits 60 models, including the non-private original-graph reference. The external phase prepares one SL cluster model, twenty reports, thirty derived graphs, and thirty models. Across the two phases, observed execution including auditing totals about 39 minutes; this batch total includes both backbones, the reference, repeated reports, and audits and is not a per-method runtime.

GLP's mean inner classifier-training times are 12.80 seconds for GPR and 2.62 seconds for GCN; Blink Soft/Hybrid GPR take 12.77/12.08 seconds, while SL's joint GCN training and iterative graph learning take 118.90 seconds. These inner-loop figures exclude loading, saving, hashing, report collection, and reconstruction, so the measurements do not establish an end-to-end speed ratio. All final models run natively on the GPU, without CPU tensor offload or reduced precision/epochs. Derived-graph caches for the two phases occupy approximately 31.86 and 43.45 GiB on disk, respectively; these are cumulative archives rather than simultaneous GPU requirements.

The current representation stores dense $n \times n$ matrices. Group reconstruction costs $O(nc^2+c^3+n^2c)$; exact all-pairs feature cosine preparation costs $O(n^2 f)$, and storage is $O(n^2 + nc)$. The full-graph Physics and Flickr extension reaches 34,493 and 89,250 nodes, respectively (Appendix C.1), using CPU reconstruction for Flickr. This expands the measured operating range but does not change the quadratic storage complexity or establish a sparse-linear implementation. There is no extra report channel compared with matched HPGR, but no general communication advantage over methods releasing only cluster counts.

The three-graph extension's 470 new checkpoints and PubMed's 90 checkpoints reproduce saved predictions exactly in the recorded execution environment. The three-graph extension's 130 newly copied reports and all 30 PubMed reports match their archived hashes. Report contents, source versions, and frozen configurations are preserved separately from isolated benchmark labels and raw-graph evaluation inputs. Bridging older graphs across environments had maximum absolute discrepancy 2.23×10^{-6}, below the prespecified 3×10^{-6} tolerance, and was not bitwise equality. Such checks support artifact consistency; they do not replace external code review or original-paper numerical calibration. Noise seeds belong to the private audit archive. The manuscript source bundle

contains only derived benchmark evaluation data needed to redraw the figures, not the private report collection archive.

C COMPLETE CLASSIFICATION RESULTS

The following four tables give the complete measured primary grid: Cora, CiteSeer, and Photo at $B = 4, 6, 8, 12$, and PubMed at $B = 4, 8$. Both Accuracy and Macro-F1 are included. Boldface identifies the best mean among private methods within a backbone and budget; the non-private original-graph reference is displayed separately. Boldface is not a significance test. The original-graph reference is trained once per repetition and reused across budget columns. Unmeasured PubMed budgets are not interpolated.

Table 4: GPR test accuracy (%, mean ± sample SD, $n = 5$) on the complete evaluated budget grid. PubMed was evaluated only at $B = 4, 8$. Bold marks the highest mean among private-report methods; the non-private original graph is a reference excluded from that comparison. Reference values are reused across budgets.

Dataset	B	GLP	HPGR	Blink Soft	Blink Hybrid	Original graph
Cora	4	**69.81 ± 1.11**	65.08 ± 0.65	61.54 ± 0.74	58.49 ± 1.84	85.61 ± 0.71
Cora	6	**73.97 ± 0.73**	69.39 ± 0.51	68.86 ± 2.04	66.12 ± 1.71	85.61 ± 0.71
Cora	8	**80.41 ± 1.32**	78.14 ± 1.18	78.55 ± 2.11	74.98 ± 0.76	85.61 ± 0.71
Cora	12	**85.54 ± 0.51**	85.29 ± 0.72	85.29 ± 0.91	85.29 ± 0.87	85.61 ± 0.71
CiteSeer	4	**70.17 ± 0.64**	68.27 ± 1.45	63.22 ± 1.81	56.88 ± 2.29	73.03 ± 0.64
CiteSeer	6	**70.94 ± 1.21**	68.82 ± 1.19	64.71 ± 1.21	58.82 ± 0.87	73.03 ± 0.64
CiteSeer	8	**69.93 ± 1.27**	68.87 ± 0.32	64.88 ± 1.52	61.18 ± 1.79	73.03 ± 0.64
CiteSeer	12	**72.05 ± 0.84**	71.71 ± 0.71	71.80 ± 0.89	71.73 ± 0.98	73.03 ± 0.64
Photo	4	**88.90 ± 0.29**	88.01 ± 0.71	88.70 ± 0.31	88.77 ± 0.63	93.55 ± 0.41
Photo	6	**91.55 ± 0.11**	91.23 ± 0.24	91.47 ± 0.20	90.53 ± 0.47	93.55 ± 0.41
Photo	8	**92.80 ± 0.44**	92.49 ± 0.58	92.54 ± 0.29	91.62 ± 0.41	93.55 ± 0.41
Photo	12	**93.77 ± 0.38**	93.51 ± 0.25	93.57 ± 0.36	93.59 ± 0.39	93.55 ± 0.41
PubMed	4	**83.28 ± 0.27**	82.47 ± 0.19	82.61 ± 0.25	81.34 ± 0.24	86.69 ± 0.15
PubMed	8	**84.31 ± 0.18**	83.35 ± 0.33	83.38 ± 0.35	83.67 ± 0.16	86.69 ± 0.15

Table 5: GPR test Macro-F1 (%, mean ± sample SD, $n = 5$) on the complete evaluated budget grid. PubMed was evaluated only at $B = 4, 8$. Bold marks the highest mean among private-report methods; the non-private original graph is a reference excluded from that comparison. Reference values are reused across budgets.

Dataset	B	GLP	HPGR	Blink Soft	Blink Hybrid	Original graph
Cora	4	**64.01 ± 1.10**	59.19 ± 1.23	55.85 ± 2.64	51.68 ± 2.71	83.68 ± 0.69
Cora	6	**69.45 ± 1.67**	64.57 ± 1.44	64.60 ± 2.40	61.73 ± 2.22	83.68 ± 0.69
Cora	8	**77.52 ± 1.81**	74.94 ± 1.66	75.35 ± 2.22	71.93 ± 0.87	83.68 ± 0.69
Cora	12	**83.59 ± 0.58**	83.22 ± 1.12	83.18 ± 0.90	83.27 ± 1.03	83.68 ± 0.69
CiteSeer	4	**64.10 ± 1.86**	63.14 ± 0.95	56.51 ± 1.00	51.18 ± 2.49	66.49 ± 0.92
CiteSeer	6	**64.72 ± 1.59**	61.91 ± 1.41	57.76 ± 2.02	53.11 ± 1.68	66.49 ± 0.92
CiteSeer	8	**63.40 ± 1.36**	63.07 ± 1.00	58.09 ± 2.29	54.99 ± 2.08	66.49 ± 0.92
CiteSeer	12	**65.26 ± 2.41**	64.99 ± 1.39	64.91 ± 1.63	65.08 ± 1.33	66.49 ± 0.92
Photo	4	85.43 ± 0.58	84.07 ± 1.13	85.11 ± 0.92	**85.60 ± 1.27**	91.89 ± 0.59
Photo	6	**89.53 ± 0.38**	88.94 ± 0.42	89.25 ± 0.65	87.61 ± 1.13	91.89 ± 0.59
Photo	8	**91.20 ± 0.62**	90.77 ± 0.80	90.67 ± 0.80	89.66 ± 0.83	91.89 ± 0.59
Photo	12	**92.32 ± 0.61**	91.99 ± 0.55	91.95 ± 0.53	92.05 ± 0.59	91.89 ± 0.59
PubMed	4	**83.23 ± 0.24**	82.34 ± 0.14	82.65 ± 0.26	80.92 ± 0.30	86.18 ± 0.15
PubMed	8	**84.26 ± 0.20**	83.25 ± 0.33	83.28 ± 0.33	83.50 ± 0.20	86.18 ± 0.15

Table 6: GCN test accuracy (%, mean ± sample SD, $n = 5$) on the complete evaluated budget grid. PubMed was evaluated only at $B = 4, 8$. Bold marks the highest mean among private-report methods; the non-private original graph is a reference excluded from that comparison. Reference values are reused across budgets.

Dataset	B	GLP	HPGR	SL-LPGL	Original graph
Cora	4	$\mathbf{69.33 \pm 1.30}$	64.37 ± 1.00	68.93 ± 1.13	84.73 ± 0.85
Cora	6	$\mathbf{73.71 \pm 1.04}$	70.25 ± 0.70	72.02 ± 0.44	84.73 ± 0.85
Cora	8	$\mathbf{79.73 \pm 0.69}$	76.81 ± 1.35	71.76 ± 0.75	84.73 ± 0.85
Cora	12	$\mathbf{84.82 \pm 1.00}$	84.54 ± 0.90	72.76 ± 0.96	84.73 ± 0.85
CiteSeer	4	$\mathbf{70.18 \pm 0.92}$	68.22 ± 1.28	69.64 ± 1.05	72.14 ± 0.48
CiteSeer	6	$\mathbf{71.29 \pm 0.81}$	69.18 ± 1.01	70.47 ± 1.04	72.14 ± 0.48
CiteSeer	8	$\mathbf{70.16 \pm 0.85}$	69.50 ± 0.79	70.05 ± 0.77	72.14 ± 0.48
CiteSeer	12	$\mathbf{71.55 \pm 0.41}$	71.16 ± 0.52	71.18 ± 0.98	72.14 ± 0.48
Photo	4	$\mathbf{88.04 \pm 0.97}$	87.23 ± 0.87	84.92 ± 0.94	92.01 ± 0.15
Photo	6	88.89 ± 3.24	$\mathbf{89.27 \pm 0.39}$	84.79 ± 1.15	92.01 ± 0.15
Photo	8	$\mathbf{91.64 \pm 0.49}$	91.35 ± 0.71	84.57 ± 1.26	92.01 ± 0.15
Photo	12	$\mathbf{92.40 \pm 0.32}$	92.19 ± 0.17	84.58 ± 1.61	92.01 ± 0.15
PubMed	4	$\mathbf{84.61 \pm 0.26}$	83.43 ± 0.19	84.07 ± 0.23	87.42 ± 0.10
PubMed	8	$\mathbf{85.58 \pm 0.24}$	84.19 ± 0.35	85.22 ± 0.24	87.42 ± 0.10

Table 7: GCN test Macro-F1 (%, mean ± sample SD, $n = 5$) on the complete evaluated budget grid. PubMed was evaluated only at $B = 4, 8$. Bold marks the highest mean among private-report methods; the non-private original graph is a reference excluded from that comparison. Reference values are reused across budgets.

Dataset	B	GLP	HPGR	SL-LPGL	Original graph
Cora	4	63.46 ± 1.14	58.59 ± 2.15	$\mathbf{63.88 \pm 0.50}$	82.86 ± 0.80
Cora	6	$\mathbf{69.32 \pm 1.97}$	65.70 ± 1.61	67.07 ± 0.92	82.86 ± 0.80
Cora	8	$\mathbf{77.12 \pm 0.67}$	73.71 ± 1.73	67.02 ± 1.34	82.86 ± 0.80
Cora	12	$\mathbf{82.79 \pm 1.00}$	82.36 ± 0.89	68.39 ± 1.27	82.86 ± 0.80
CiteSeer	4	$\mathbf{64.94 \pm 1.01}$	63.31 ± 1.39	62.70 ± 1.27	66.02 ± 0.75
CiteSeer	6	$\mathbf{64.26 \pm 0.60}$	62.66 ± 1.52	64.10 ± 1.07	66.02 ± 0.75
CiteSeer	8	64.17 ± 0.79	63.67 ± 0.86	$\mathbf{65.10 \pm 0.90}$	66.02 ± 0.75
CiteSeer	12	$\mathbf{65.87 \pm 0.51}$	64.89 ± 0.72	65.18 ± 0.94	66.02 ± 0.75
Photo	4	$\mathbf{83.81 \pm 1.66}$	82.41 ± 1.57	75.94 ± 1.25	89.89 ± 0.32
Photo	6	83.07 ± 8.41	$\mathbf{86.82 \pm 0.57}$	75.69 ± 1.71	89.89 ± 0.32
Photo	8	$\mathbf{89.90 \pm 0.57}$	89.41 ± 1.33	75.17 ± 1.69	89.89 ± 0.32
Photo	12	$\mathbf{90.57 \pm 0.51}$	90.01 ± 0.09	75.81 ± 2.22	89.89 ± 0.32
PubMed	4	$\mathbf{84.54 \pm 0.26}$	83.34 ± 0.18	83.98 ± 0.20	86.86 ± 0.13
PubMed	8	$\mathbf{85.46 \pm 0.22}$	84.09 ± 0.33	85.12 ± 0.22	86.86 ± 0.13

PubMed adds a larger, sparse benchmark to the utility comparison. At $B = 4$, GLP improves over matched HPGR by 0.81 pp with GPR and 1.18 pp with GCN; at $B = 8$, the corresponding gains are 0.97 and 1.38 pp. GLP also outperforms Blink Soft/Hybrid under GPR and SL-LPGL under GCN at both budgets, with margins of 0.36–1.94 pp. The margins are smaller than on the citation graphs at low budget, so the larger graph establishes an additional operating condition at 19,717 nodes rather than a universal ranking.

C.1 Full-graph Physics and Flickr evaluation

The extension evaluates the entire Coauthor Physics graph (34,493 nodes, 8,415 features, five classes) and Flickr graph (89,250 nodes, 500 features, seven classes). No nodes or features are subsampled, and the propagation matrices retain the original dense definition. The public cosine neighborhood uses raw attributes and $k = 10$; $\alpha = 0.5$, the capacity rule, the privacy allocation, and classifier parameters are unchanged.

Partitions and repetition counts. Both graphs use sequential stratification with target proportions 10/25/25/40% for training, validation, test, and unused labels. Physics has 3,449/8,623/8,623/13,798 nodes in these sets; Flickr has 8,925/22,312/22,312/35,701. The initial split seeds are 2026091901 and 2026091902, respectively. These are controlled partitions; in particular, Flickr does not use its

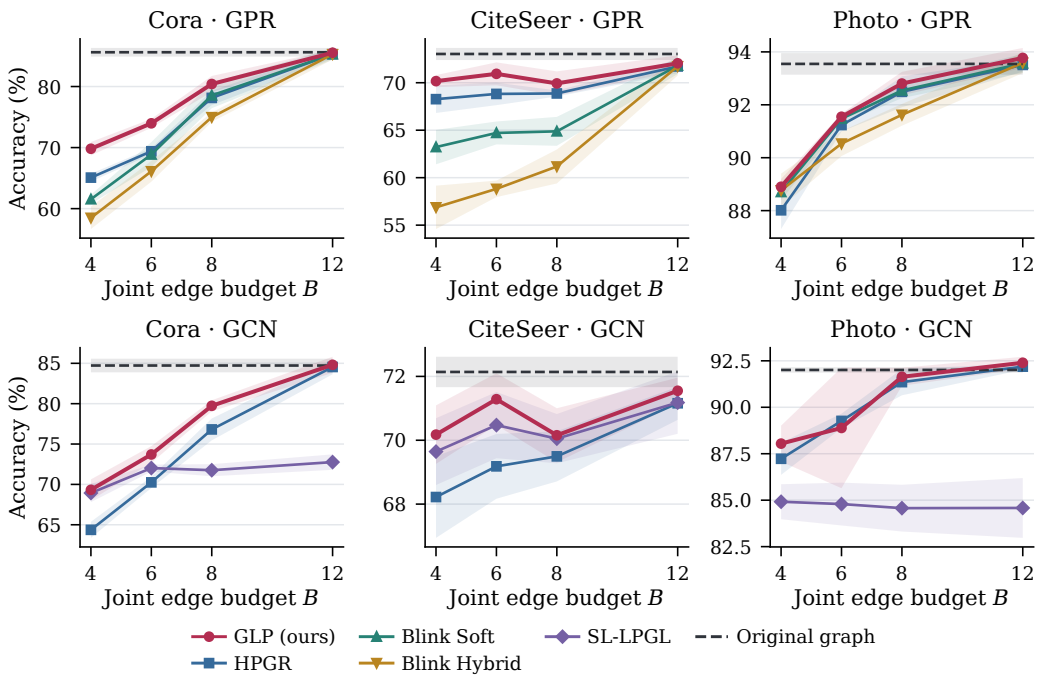


Figure 5: Accuracy across the complete budget grid on three graphs, with GPR and GCN shown separately. All methods show five-run means and sample SDs. The original-graph reference uses no private report and is repeated across budgets. PubMed's two measured budgets appear in Table 1 and Figure 6.

official split, so these numbers should not be directly ranked against results obtained on that split. Test labels are isolated from reconstruction, training, and checkpoint selection.

Physics evaluates $B = 4, 8$ with five report/classifier repetitions. Flickr evaluates $B = 4$ with three repetitions, not five, and no $B = 8$ result is claimed. The feature-only membership models are fixed, with seeds 28001 (Physics) and 29001 (Flickr); classifier seeds are 28010–28014 and 29010–29012. Each repetition uses a newly saved report; GLP and HPGR share that report and its classifier seed. These repetitions quantify report and training variability on a fixed partition, not variability across independent data splits. All SDs are sample SDs, and boldface denotes the highest measured mean rather than statistical significance.

Coverage and reference conditions. Physics contains 80 classifier results: 40 for GLP/HPGR (two backbones, two budgets, five repetitions), 20 for Blink Soft/Hybrid (GPR, two budgets, five repetitions), and 20 for public-feature/original-graph references (two backbones, five repetitions). Flickr contains 24 results: GLP, HPGR, public-feature, and original-graph conditions under both backbones, each repeated three times. Blink retains its own query mechanism and shared Soft/Hybrid reports. The public-feature and original-graph operators are reused across budgets or repetitions where unchanged; classifiers retain their prescribed seeds.

Public W is a feature-only reference using allowed public information. The original graph uses unrandomized benchmark edges and is a non-private reference; neither is treated as a private-report baseline in the highlighted rankings. Physics SL-LPGL training was deferred after its initializer exceeded GPU memory and then succeeded with a memory-allocation adjustment. Flickr Blink and SL-LPGL were not evaluated. Missing conditions are not zero scores, and the large-graph results establish only the comparisons actually reported here.

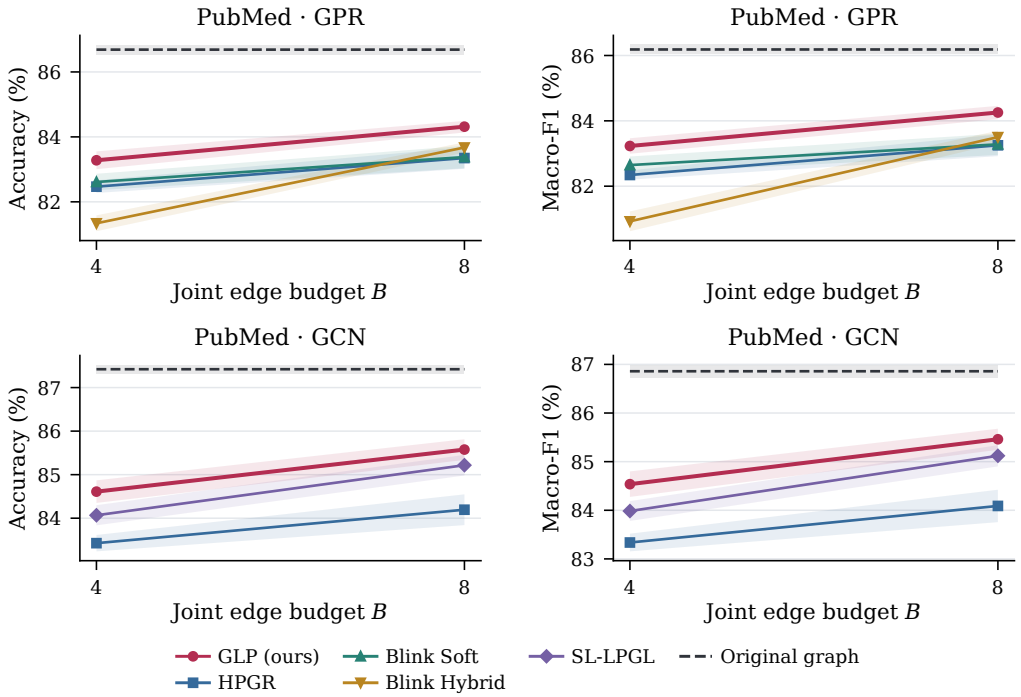


Figure 6: Complete PubMed comparison at the two measured budgets. GPR and GCN are shown separately for Accuracy and Macro-F1; every method shows five-run means and sample SDs. The same original-graph reference run is reused at both budgets.

Table 8: Complete Physics results (%, mean ± sample SD; 5 repetitions). Bold selects among evaluated private methods within each backbone, budget, and metric. Public W and the original graph are reference conditions, separated below each private block.

Method	B	Accuracy	Macro-F1
GPR			
GLP	4	**95.55 ± 0.08**	**94.23 ± 0.08**
HPGR	4	94.72 ± 0.11	93.13 ± 0.16
Blink Soft	4	94.76 ± 0.24	93.26 ± 0.34
Blink Hybrid	4	94.15 ± 0.21	92.67 ± 0.25
GLP	8	**95.96 ± 0.15**	**94.74 ± 0.22**
HPGR	8	95.51 ± 0.08	94.15 ± 0.12
Blink Soft	8	95.49 ± 0.11	94.20 ± 0.18
Blink Hybrid	8	94.61 ± 0.13	93.11 ± 0.20
Public W	—	96.02 ± 0.04	94.80 ± 0.06
Original graph	—	95.82 ± 0.09	94.45 ± 0.13
GCN			
GLP	4	**95.45 ± 0.11**	**94.09 ± 0.15**
HPGR	4	94.60 ± 0.08	92.98 ± 0.13
GLP	8	**95.89 ± 0.08**	**94.69 ± 0.11**
HPGR	8	95.35 ± 0.07	93.92 ± 0.11
Public W	—	95.92 ± 0.05	94.68 ± 0.08
Original graph	—	95.82 ± 0.08	94.51 ± 0.11

Table 9: Complete Flickr results (%, mean ± sample SD; 3 repetitions). Bold selects among evaluated private methods within each backbone, budget, and metric. Public W and the original graph are reference conditions, separated below each private block.

Method	B	Accuracy	Macro-F1
GPR			
GLP	4	45.56 ± 0.29	$\mathbf{14.92 \pm 0.63}$
HPGR	4	$\mathbf{45.77 \pm 0.18}$	14.63 ± 0.30
Public W	—	45.73 ± 0.18	14.99 ± 0.27
Original graph	—	51.00 ± 0.27	20.95 ± 0.57
GCN			
GLP	4	$\mathbf{44.03 \pm 0.36}$	$\mathbf{14.19 \pm 0.35}$
HPGR	4	42.89 ± 0.12	12.03 ± 0.35
Public W	—	45.52 ± 0.08	14.64 ± 0.24
Original graph	—	50.32 ± 0.22	18.74 ± 1.90

Paired gains and the public-information boundary. On Physics, all 20 paired GLP–HPGR accuracy differences are positive. GPR improves by 0.83/0.45 pp and GCN by 0.85/0.55 pp at $B = 4/8$, respectively. Macro-F1 improves by 1.11/0.59 pp for GPR and 1.11/0.77 pp for GCN. Against Blink Soft, GPR accuracy gains are 0.79/0.47 pp; against Blink Hybrid they are 1.40/1.35 pp. At $B = 8$, GLP's accuracy is within 0.06 pp of Public W under both backbones, while its mean is still slightly lower. These results support gains over evaluated private reconstructions, not a universal advantage over public-only learning.

On Flickr, all three GCN accuracy differences favor GLP, whereas all three GPR differences favor HPGR. The GCN mean gains are 1.13 pp in accuracy and 2.16 pp in Macro-F1; the GPR changes are -0.21 and $+0.30$ pp. Public W remains stronger in accuracy and Macro-F1 under both backbones, and the original graph is substantially stronger. Thus successful full-graph execution and gains over a matched private reconstruction do not establish that privatized topology is necessary or sufficient when public features already provide a competitive operator. The differing metric response also cautions against using accuracy alone to summarize the Flickr comparison.

Execution and memory. The extension uses a separate Linux server with approximately 566 GiB host memory, Python 3.12.14 and PyTorch 2.11.0+cu128. Each selected GPU exposes approximately 48 GiB to the runtime; a stage uses one GPU rather than pooling memory across devices. Physics reconstruction and training run on GPU. Flickr reconstruction runs on CPU, followed by GPU training, preserving the dense method instead of substituting a sparse approximation.

Table 10: Recorded first-run Flickr costs for GLP at $B = 4$. Wall time covers the profiled stage, including its loading and other stage overhead; CPU RSS and allocated GPU memory are per-stage peaks. These are single-run measurements, not averages over all repeats or an end-to-end speed comparison.

Stage	Wall time (s)	CPU peak (GiB)	GPU peak (GiB)
Dense reconstruction	1,116.2	223.36	—
GPR training stage	436.0	38.66	30.07
GCN training stage	150.4	38.59	29.88

Table 10 records feasibility at 89,250 nodes and the cost of the current implementation. Reconstruction takes about 18.6 minutes, but requires substantially more host memory than the smaller-graph experiments. The all-pairs similarity and dense probability operators still require quadratic storage; these measurements do not demonstrate linear scaling or feasibility on million-node graphs. Report collection is a separate stage and is not included in the reconstruction timing. Memory peaks cannot be added across stages to estimate simultaneous use, and stage wall times are not the classifier-only times reported for PubMed.

C.2 New complete-split confirmation

Tables 11 and 12 supplement Figure 7. The new splits support repeatable improvements over the group-prior baseline under frozen parameters. Paired changes use each saved report and its matched model seed. All eight mean changes are positive for both metrics.

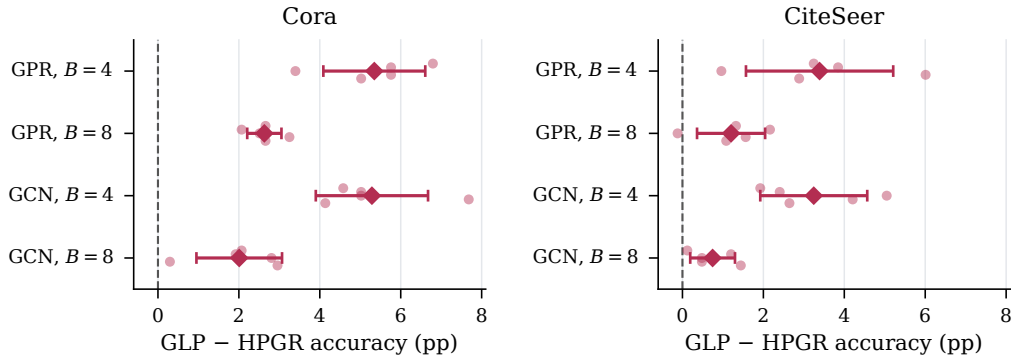


Figure 7: New complete-partition confirmation: paired accuracy changes relative to matched HPGR. Points show individual runs; summaries show means and sample SDs. Method parameters were frozen before the new partitions and reports were generated.

Table 11: New complete-split confirmation: test accuracy (%, mean ± sample SD, $n = 5$). Splits, pseudo-label models, and reports are newly constructed with frozen method parameters on the same two graphs. Bold compares GLP and HPGR; the original graph is a reference.

Dataset	Model	B	GLP	HPGR	Original graph
Cora	GPR	4	**70.31 ± 1.37**	64.96 ± 1.79	84.49 ± 0.28
Cora	GPR	8	**80.53 ± 0.89**	77.90 ± 0.79	84.49 ± 0.28
Cora	GCN	4	**70.43 ± 1.06**	65.14 ± 0.72	84.46 ± 0.91
Cora	GCN	8	**80.21 ± 0.76**	78.20 ± 0.56	84.46 ± 0.91
CiteSeer	GPR	4	**68.10 ± 0.85**	64.71 ± 1.09	72.50 ± 0.51
CiteSeer	GPR	8	**68.75 ± 0.74**	67.55 ± 0.96	72.50 ± 0.51
CiteSeer	GCN	4	**66.61 ± 1.10**	63.37 ± 1.28	70.55 ± 0.47
CiteSeer	GCN	8	**67.69 ± 0.69**	66.95 ± 0.99	70.55 ± 0.47

Table 12: New complete-split confirmation: test Macro-F1 (%, mean ± sample SD, $n = 5$). Splits, pseudo-label models, and reports are newly constructed with frozen method parameters on the same two graphs. Bold compares GLP and HPGR; the original graph is a reference.

Dataset	Model	B	GLP	HPGR	Original graph
Cora	GPR	4	**67.07 ± 1.71**	61.31 ± 3.03	83.20 ± 0.43
Cora	GPR	8	**79.10 ± 1.26**	76.22 ± 1.00	83.20 ± 0.43
Cora	GCN	4	**67.47 ± 1.10**	61.64 ± 0.65	83.22 ± 1.09
Cora	GCN	8	**78.84 ± 0.77**	76.57 ± 0.79	83.22 ± 1.09
CiteSeer	GPR	4	**62.66 ± 1.05**	58.88 ± 2.55	66.62 ± 0.47
CiteSeer	GPR	8	**62.51 ± 1.95**	62.11 ± 0.87	66.62 ± 0.47
CiteSeer	GCN	4	**62.37 ± 0.96**	59.09 ± 1.61	65.57 ± 0.46
CiteSeer	GCN	8	**63.24 ± 0.64**	62.26 ± 0.76	65.57 ± 0.46

D MECHANISM AND BOUNDARY ANALYSES

D.1 UNCERTAINTY, CORRECTION OPPORTUNITIES, AND BREAKAGE

Membership entropy is $h_i = -\sum_a C_{ia} \log C_{ia} / \log c$, with $0 \log 0 = 0$. Tertile thresholds are computed using all nodes' allowed membership vectors; test labels only score the resulting strata. In each stratum and repetition, let e_0 be the baseline error fraction, r_{fix} the fraction of those errors corrected by GLP, and r_{break} the fraction of baseline-correct predictions made wrong. The exact counting identity is

$$\Delta\text{Acc} = e_0 r_{\text{fix}} - (1 - e_0) r_{\text{break}}. \tag{16}$$

Figure 8 shows these diagnostics on the new splits. We average per-run ratios and products. Multiplying their separately averaged ratios is not generally equivalent. These statistics describe overlapping nodes in repeated runs; they do not provide independent node-level significance tests.

On the new GPR splits at $B = 4$, mean net accuracy gains in the low/middle/high-entropy thirds are 0.11/4.34/10.33 pp for Cora and 0.64/0.68/7.80 pp for CiteSeer. Cora's high-entropy group has a 31.37% correction rate and a 12.69% breakage rate, compared with 4.00% and 0.36% in its low-entropy group. Greater uncertainty accompanies both correction opportunities and damage to previously correct predictions; it does not by itself establish an adaptive mixing rule.

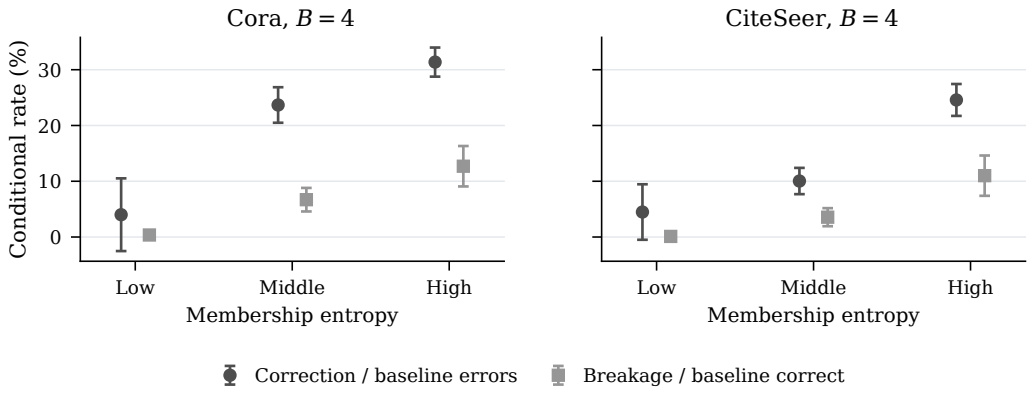


Figure 8: New complete-split GPR at $B = 4$: conditional correction and breakage rates by public-membership entropy. Denominators are baseline-wrong and baseline-correct nodes, respectively. Summaries show five-run means and sample SDs. These are post hoc explanations, not prospective selectors.

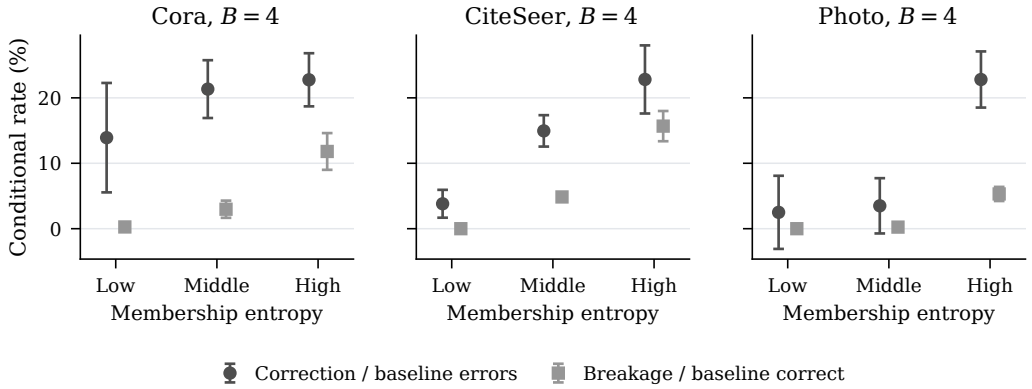


Figure 9: Original-split GPR at $B = 4$: conditional correction and breakage rates by public-membership entropy. Summaries show five-run means and sample SDs. Both rates rise with entropy on all three graphs in this condition. The trend does not establish that entropy itself causes benefit or that it should directly set a larger mixing weight.

D.2 Structure recovery and task performance

For each saved edge-probability matrix P, whole-graph Brier error averages $(P_{ij} - A_{ij})^2$ over all ordered non-self positions. Edge Brier restricts this average to true edges; nonedge Brier restricts it to nonedges. These are diagnostic uses of the isolated benchmark graph, not inputs to the method. The normalized membership-propagation error is

$$E_C(P) = \frac{\|(S(P) - S(A))C\|_F}{\|S(A)C\|_F}. \tag{17}$$

Figure 4 reports each structural metric's mean per-run relative reduction $100(E_{\mathrm{HPGR}} - E_{\mathrm{GLP}})/E_{\mathrm{HPGR}}$, separately from the paired accuracy change in pp. A positive ordinate denotes improvement, but the two units should not be compared numerically.

Across the citation graphs' four original and two confirmatory budgets, all 60 paired values improve for whole-graph Brier, edge Brier, and E_C. Photo's E_C instead worsens in all 20 pairs. Its whole-graph Brier worsens in all 15 pairs at $B = 4, 6, 8$ and improves in the five pairs at $B = 12$. At $B = 4$, mean E_C rises from 0.354 to 0.421 while GPR improves by 0.89 pp; at $B = 6$, it rises from 0.247 to 0.333 while GPR improves by 0.32 pp. This does not imply that worse recovery causes better classification. It establishes that these aggregate proxies are insufficient to predict the direction of task change.

D.3 Fixed-model propagation diagnostics

An earlier saved-model diagnostic uses the original Cora/CiteSeer splits, $B = 4, 8$, and five reports per condition. It freezes both the baseline and GLP GCN checkpoints and substitutes the two propagation graphs while computing only allowed training/validation cross-entropy. At the baseline parameters and $\alpha = 0$, the direct-edge and degree-normalization terms in equation (9) give Table 13. These are loss derivatives with respect to α, not accuracy increments or predictions for a half-unit step.

Table 13: Mean validation-loss derivative over five saved-model diagnostics. Sign changes compare the direct-edge term against the full derivative.

Dataset	B	Direct edge	Degree terms	Full	Sign changes
Cora	4	−1.546	−4.062	−5.608	0/5
Cora	8	−23.692	−14.701	−38.394	0/5
CiteSeer	4	+0.852	−7.072	−6.220	4/5
CiteSeer	8	+34.399	−95.835	−61.436	5/5

Table 15: Historical Cora injection-location controls with GPR (test accuracy in %, mean ± sample SD, $n = 5$). Parentheses give the historically selected α. Each family and training-label fraction selected from $\{0, 0.1, 0.5, 1\}$ using five-run mean validation accuracy, with smaller α breaking ties; validation labels occupy an additional 25% of nodes. The same reports are reused within each fraction. Bold marks the highest mean in each column; these exploratory controls preceded the frozen confirmation.

Method	50% training	10% training	5% training
HPGR	83.46 ± 0.42 (0)	77.75 ± 1.71 (0)	74.71 ± 1.92 (0)
Uniform prior + RR	83.81 ± 0.43 (0.1)	78.20 ± 1.33 (0.1)	74.89 ± 1.59 (0.1)
Local prior + RR	$\mathbf{84.58 \pm 0.61}$ (0.5)	$\mathbf{79.44 \pm 1.87}$ (0.5)	$\mathbf{76.04 \pm 2.20}$ (0.5)
Propagation mixture	83.49 ± 0.66 (0.1)	78.35 ± 1.60 (0.1)	73.23 ± 1.79 (0.5)

With baseline parameters fixed, substituting the GLP graph decreases validation cross-entropy in all 20 cases; this also holds in all 20 cases with GLP parameters fixed. Yet the baseline-parameter substitution increases training cross-entropy in all 20 cases. This is consistent with a conditional fit–validation tradeoff, not proof of a particular regularization mechanism. The complete local direction is favorable in all cases, whereas retrained test accuracy improves in only 18/20. Omitting the RR multiplier in a diagnostic also gives a favorable sign in all 20 cases, so the batch does not show that the full score is a better selector. Checkpoints were already selected on the same validation set; this remains exploratory conditional evidence.

Photo's extension provides a complementary boundary. At $B = 6$ with GCN, GLP reaches $88.89 \pm 3.24\%$ accuracy against HPGR's $89.27 \pm 0.39\%$; the paired mean difference is -0.38 pp and 2 of the 5 paired changes are negative. Checkpoint replay confirms these values. Mean validation cross-entropy is 0.3518 for baseline parameters on the baseline graph and 0.4543 on the GLP graph. For GLP parameters, the corresponding values are 0.3816 and 0.3964. The worst GLP run still has validation cross-entropy 0.5180 after switching back to the baseline graph, above the original baseline's 0.3221. Thus learned parameters as well as the inference operator participate in the gap. The experiment does not isolate a unique optimization failure, oversmoothing mechanism, or causal percentage attributable to each component.

D.4 Capacity and within-support weights

On Photo, inferred mass exceeds the capped top-10 support capacity, so the frozen construction retains support locations but saturates relative weights. To test whether those weights contain additional information, two controls use identical support, total mass, support-local mass $0.5K$, and background: one keeps reweighted public similarities and the other is binary within support. Table 14 reports their results. The weighted control improves over the binary control by 0.32/0.08 pp at $B = 4/8$, with four of five positive paired changes in each condition. Neither mean exceeds frozen GLP. The clean weight comparison is between these two controls; their comparison to GLP also changes local mass from K to $0.5K$ and cannot isolate a weight effect.

Table 14: Photo capacity control with GPR: accuracy (%, mean ± sample SD, $n = 5$). Weighted and binary controls share half-support local mass and background; both differ in local mass from frozen GLP. Bold marks the highest mean in this control.

B	GLP	Weighted half-mass	Binary half-mass
4	$\mathbf{88.90 \pm 0.29}$	88.85 ± 0.57	88.53 ± 0.47
8	$\mathbf{92.80 \pm 0.44}$	92.55 ± 0.42	92.47 ± 0.48

E Additional design and component controls

E.1 Injection position and the original mixing-weight selection

The earlier Cora study at $B = 8$ compared three families at nested 50%/10%/5% training-label fractions, with five repetitions each. Validation/test partitions stayed fixed, with approximately 25%

validation labels. Each family selected from $\{0, 0.1, 0.5, 1\}$ by mean validation accuracy, breaking ties toward the smaller weight. Table 15 reports selected test scores; Figure 10 shows the measured validation grid.

The local family pools $(1-\alpha)Q_g + \alpha T$ before the saved RR update. The uniform family replaces T with $\rho(J - I)$, preserving prior mass while removing the local preference. The propagation family instead uses

$$S_{\text{mix}} = (1 - \alpha)S(P_{\text{HPGR}}) + \alpha S(W),$$

after normalizing each graph with self-loops. The propagation endpoint $\alpha = 1$ applies the normalized public-feature operator directly, without report-derived mass; the local endpoint still uses report-derived mass and RR responses. These particular mixtures do not exhaust possible injection schemes.

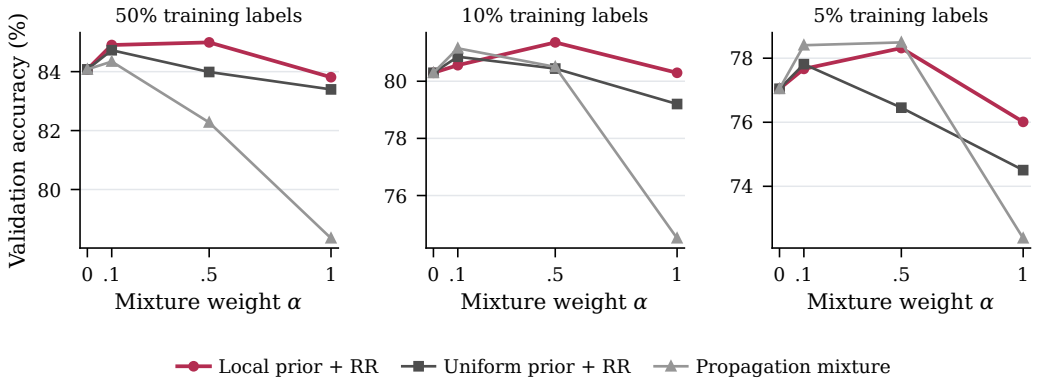


Figure 10: Original Cora validation selection at $B = 8$: five-run means at measured weights $\alpha \in \{0, 0.1, 0.5, 1\}$, without uncertainty bars. All three families use the same candidate set and selection rule; test scores do not select candidates.

The local family selects 0.5 and yields the largest selected test mean at all fractions; uniform smoothing selects 0.1, and propagation mixing selects 0.1/0.1/0.5 at 50%/10%/5%. At 5%, propagation mixing with $\alpha = 0.5$ has higher validation accuracy than 0.1 (78.49% versus 78.40%) but lower test accuracy (73.23% versus 75.36%); we retain the validation-selected result. Subsequent datasets and complete-split confirmation retain $\alpha = 0.5$. This limited grid documents selection, not a broad cross-dataset stability plateau or optimal adaptive rule.

E.2 Matched-amplitude and information-component controls

These controls use frozen GPR and $\alpha = 0.5$, with three overlapping training subsets at each of 10% and 5% training labels and five repetitions per subset. Validation/test sets remain fixed, with approximately 25% validation labels. The 15 observations per cell are not independent splits and differ from complete-partition confirmation. Both citation graphs satisfy $M < K$, where the capacity construction equals the historical target; the archived audit reproduced all 120 historical propagation graphs exactly. These controls therefore evaluate the same method. Baseline and GLP runs reused by the primary grid are not counted again.

Relationship information beyond prior mass and amplitude. At $B = 8$, we permuted the upper-triangular entries of T and mirrored them to obtain T^π, preserving the target's entry multiset and total mass. The control prior is

$$Q_\pi = (1 - a_\pi)Q_g + a_\pi T^\pi, \qquad a_\pi = \frac{\|0.5(T - Q_g)\|_F}{\|T^\pi - Q_g\|_F}.$$

All a_π were feasible (0.4951–0.4973), without validation adjustment. Prior mass and Frobenius change relative to Q_g are matched before the same saved RR update. Table 16 shows positive gains

in all four setting means and all 12 subset means, with 57/60 pairs positive; gains over HPGR remain at 5% training labels. Unlike node permutation, entry permutation need not preserve topology or spectrum. Matched prior amplitude also need not imply matched posterior amplitude after nonlinear RR.

Table 16: Earlier $B = 8$ GPR controls: accuracy (%) and paired gain (pp), mean $\pm$ sample SD over three training subsets $\times$ five repetitions. "Matched" denotes entry permutation with matched prior mass and change norm.

Dataset	Train	HPGR	GLP	Matched	GLP − matched
Cora	10%	77.89 ± 1.62	79.44 ± 1.50	78.25 ± 1.32	+1.19 ± 1.09
Cora	5%	73.31 ± 2.04	75.38 ± 1.31	73.14 ± 1.92	+2.25 ± 1.41
CiteSeer	10%	69.20 ± 0.77	70.38 ± 1.09	68.45 ± 0.94	+1.94 ± 1.47
CiteSeer	5%	66.15 ± 1.74	67.00 ± 1.33	64.80 ± 1.25	+2.20 ± 0.81

Group structure and the RR component. The density control replaces Q_g in the pooled prior by a uniform off-diagonal matrix with entry ρ, retaining T and the same RR responses. It retains private mass information but removes the heterogeneous group-prior pattern. The no-RR control propagates on $Q_{0.5}$ directly. Table 17 reports GLP minus each control. Group structure improves all eight means, with 58/60 positive pairs at each budget. The RR contribution is strongly beneficial on Cora at $B = 8$ but negative on CiteSeer at both budgets, supporting a conditional task-utility interpretation. Each within-budget comparison reuses the same report; omitting a component does not refund its released privacy budget. Across budgets, both group and RR noise change, so the differences do not identify a single-channel causal effect.

Table 17: Component contributions in paired accuracy points (mean $\pm$ sample SD, 15 observations per cell). Group: GLP minus density control. RR: GLP minus no-RR control. Negative RR effects are retained.

		Group contribution		RR contribution	
Dataset	Train	$B = 4$	$B = 8$	$B = 4$	$B = 8$
Cora	10%	+1.25 ± 1.02	+1.66 ± 1.08	+0.80 ± 1.26	+10.54 ± 1.24
Cora	5%	+3.58 ± 1.73	+2.56 ± 1.49	+0.69 ± 1.29	+10.39 ± 1.74
CiteSeer	10%	+2.26 ± 1.04	+4.13 ± 1.32	−1.52 ± 1.06	−1.87 ± 1.70
CiteSeer	5%	+4.42 ± 2.27	+6.21 ± 1.54	−2.37 ± 1.71	−3.44 ± 1.80